SEMI-ANNUAL PROGRESS REPORT

Period: January-June 1994

"The Effects of Cloud Inhomogeneities Upon Radiative Fluxes, and the Supply of a Cloud Truth Validation Dataset"

Submitted by

Institute of Atmospheric Sciences
South Dakota School of Mines and Technology
501 E. St. Joseph Street
Rapid City, SD 57701-3995

to

National Aeronautics and Space Administration
Goddard Space Flight Center
Contract NAS5-31718
Greenbelt, MD 20771

1. Work Accomplished During the Report Period

This work is directed towards the development of algorithms for the ASTER and HIRIS science/instrument teams. Special emphasis is being placed on a wide variety of cloud optical property retrievals, and especially retrievals of cloud and surface properties in the polar regions.

2. Research Activities

2.1 Cloud algorithms

2.1.1 ASTER Polar Cloud Mask

During this reporting period, the first draft of the Algorithm Theoretical Basis Document (ATBD) for the ASTER Polar Cloud Mask was reviewed by the ASTER Science Team. In addition, on 18 February, Dr. Ron Welch briefed the Science Team on the salient aspects of the document. Suggestions and comments were provided by the Science Team, which were subsequently incorporated into a rewrite of the document. The new version was resubmitted and was forwarded to the EOS program office for a peer review. In May, Ron Welch attended the ATBD Review and presented

an overview of the algorithm for the Polar Cloud Mask Standard Product, in addition to the algorithms for the Cloud Special Products, and the Sea Ice Special Products. He also attended the joint Japanese-American ASTER Team Meeting. Rand Feind attended the Third Circumpolar Symposium on Remote Sensing of Arctic Environments held the week of 14 May in Fairbanks, Alaska and presented the paper "The ASTER Polar Cloud Mask Algorithm."

The overall structure of the algorithm described in the ATBD is multistage or hierarchical. The first stage classifies the unambiguous feature vectors (i.e., the ones that are located close to class cluster centers) in a fast process using table lookups. For example, large contiguous areas (i.e., hundreds of thousands to a few million pixels) of water or thick water cloud, can be classified in less than a few minutes. Only spectral features are utilized in this stage. The remaining unclassified, ambiguous pixels are then assigned probable class memberships based on the proximity of their feature vector. They then are passed on to a fuzzy expert. Here textural features are computed to augment classification. The fuzzy expert applies rules based on past experience to obtain a certainty measure associated with a classification. In this stage, potentially 2 classes are associated with each pixel if the fuzzy expert determines the pixel belongs to a mixed class. The fuzzy expert is also augmented by the probable class memberships from the first stage and spatial context. The spatial context is determined by the classes of the nearest neighbor pixels. At this point all pixels in the scene are classified (including an unknown class). As a final stage, a quality assurance test is performed. Pixels are selected at random and reclassified using an artificial neural network. The statistics for classification agreement are then used to assess the confidence in the classification. A copy of the current version of the ATBD is attached.

Our 23 Landsat TM Antarctic polar scenes were loaded onto disk and we have selected approximately 200 training samples from those scenes. An eigenvector analysis was conducted on those training samples and the results indicated that water, snow/ice, and thick cloud samples are well separated. This eigenvector analysis included the use of band ratios and differences; however, no textures as yet. These results are encouraging in that it appears significant portions of our polar scenes can be classified with only spectral features using fast classification techniques. During the next reporting period we will be generating the textural features for our samples and also be selecting additional samples for classes in which we are deficient or for classes which appear to be causing problems for the classifier. The first version of the fast classifier is being tested on these scenes.

Because our aforementioned data set was limited to Antarctic summertime scenes, during this reporting period we examined several hundred Landsat TM Arctic scenes available from the EOSAT archive and ordered 11 additional scenes. The imagery selected is of areas in Alaska, Greenland, and Iceland. Vegetation, mountains, rivers, lakes and melt ponds are present in these scenes - features not found in our Antarctic data set. The additional data (44 9 track tapes total), each containing a quad of Landsat TM data, were received during April and were loaded onto the system. An additional 100 training samples have been extracted from this data and the first version of the classifier is being tested on them. The classifier is being modified to accommodate the new classes found in these scenes. The training samples obtained indicate that the ratio of TM Bands 4 and 2 are very useful for identifying land features. An algorithm that works well on both the Antarctic and Arctic data sets should be fairly robust.

The current version of the algorithm (based only on spectral features and includes band ratios and differences) is being tested on our 22 Landsat TM Antarctic scenes and 44 Landsat TM Arctic scenes. It appears that at least 60 percent of the pixels in these scenes can be classified with a high degree of confidence using only the spectral features and using fast algorithms based on table lookups and simple thresholding. It also appears that the classification of a large fraction of the remaining pixels can be narrowed to 2 or 3 choices with a high degree of confidence, again using only spectral features. For example, it is possible to determine that a spectral feature belongs to the class of water or thin cloud over water, but not to any other class. We have also found that the use of low pass filtering results in additional unambiguous feature vectors. As expected the regions of thin ice cloud cause most of the problems for the classifier, especially when they are over snow or ice covered surfaces. They are equally bright in the visible bands and are equally dark in the near IR bands. Temperature is not a reliable indicator either. We are still hopeful that the bands differences in the thermal IR bands of ASTER will help resolve these thin ice clouds due to their different emissivities. Unfortunately we currently do not have any multispectral thermal IR imagery available. We expect that some MAS data obtained over the Beaufort Sea will become available in the upcoming months.

Tests of more traditional classification techniques such as Euclidean distance and Mahalanobis indicate that they provide good classification results when applied on a scene by scene basis (i.e., using the training statistics unique to each scene), but perform poorly when applied over the entire data set (i.e., using the combined statistics for all the scenes).

During the next reporting period, we will assessing the usefulness of textures in classification. The work of previous investigators indicates some

textural measures are more useful than others for classification; however, most of those studies were conducted with relatively low resolution imagery (e.g., AVHRR). Since our algorithm will operate on high spatial resolution imagery, we will be investigating the usefulness of a large set of textures, in the hopes of discovering a texture that is useful in the classification, that previous investigators have not utilized. Of course, eventually, the final feature vector will be pruned down to less than 10 elements. We will also be investigating the optimum distance parameter for the textures based on the gray level distance vector. The distance parameter used for computing the gray level difference vectors for the AVHRR imagery are not necessarily applicable in our high spatial resolution Landsat TM imagery. Wavelet transforms also will be applied to these scenes to determine if they can be useful in resolving ambiguous feature vectors. These transforms are a texture measure that not only provide an indication of the spatial frequencies present in the imagery but also the spatial location of those frequencies.

Timing tests were also conducted between C and IDL for various types of scalar, vector, and matrix operations. These timing tests will aid us in determining at what point in the algorithm development process, our IDL routines should be converted to C. (In addition, if there is any possibility of having IDL functionality available through the toolkit, these results might provide the rationale for making it available.) In some cases, C operations are superior and, in others, IDL is. For example, C is much faster in performing sequential scalar operations, while IDL is much faster in performing vector and matrix operations. However, IDL is only faster in performing vector and matrix operations when memory only needs to be allocated once initially. If memory allocation is required prior to each operation, then C is superior.

In a parallel effort, we are also investigating the viability of using a relatively new technique called multistage neural networks. They are based on a hierarchical structure in which a set of relatively small specialized neural networks are connected together by a switching network. The switching network selects a unique feature vector to be input to one of the specialized neural networks. The switches select the feature vector and the appropriate network based on expert knowledge of the classification process. This system has several advantages over a single neural network. Expert knowledge or experience can be implemented in the structure. It is potentially faster and more accurate. It can also be trained faster as each of the individual networks can be trained in parallel. If results indicate that this system is faster and/or more accurate than the fuzzy expert, then it will supplant the fuzzy expert in the second stage of the algorithm and the fuzzy expert will be used to perform the quality assurance test.

The preprocessing module for line/column dropout detection and reconstruction is nearly complete. This first version uses a simple threshold on the mean line/column gray level for missing line/column detection. Three types of reconstruction techniques are being incorporated and are selectable. They are nearest neighbor, linear interpolation, and inter-band cross correlation. Only single and double line/column dropouts are reconstructed. Three or more consecutive line/column dropouts are not reconstructed and are labeled as bad data. A simple statistic for the number of line/column dropouts for the entire scene is also computed. If it exceeds a threshold (to be determined), the scene is labeled as unsuitable for cloud masking. We are also investigating techniques for detecting and correcting striping in the imagery. We do not have any good samples of striped imagery so we are initially simulating the problem. We are testing the FFT as a tool for detecting the striping in the imagery. We are planning to test frequency domain filtering and Finite Impulse Response filters for destriping the imagery.

2.1.2 Monte Carlo Simulation of 3D Cloud Effects

We ran some simple plane parallel test cases for some highly peaked phase functions and verified that the radiance pattern obtained from the Monte Carlo simulation matches that obtained from an analytical model. We are currently modifying the model to simulate cloud shapes approximated by hyperboloids of 2 sheets. The rationale for using hyperboloids of 2 sheets stems from the work of Kuo *et al.*, 1993 in which they showed that maritime cumulus cloud fields are best fit by this geometric shape. During the next reporting period we will be simulating the radiance pattern of cloud fields in which the underlying surface reflectance is based on the bidirectional reflectance functions of Hapke. We also plan to model various types of cloud fields and compare the radiance pattern to those obtained from plane parallel results. We hope to determine the cloud optical thickness and effective radius retrieval errors when using the plane parallel assumption on cellular cloud fields.

2.1.3 Registration of TIMS to AVIRIS Imagery Using Cloud Morphology

A review of the paper "Registering TIMS to AVIRIS Imagery Using Cloud Morphology" was received from the editor of IEEE Geoscience and Remote Sensing. The recommended changes were applied and the paper was resubmitted. We expect that the paper will be accepted for publication and a final version of the paper will be provided as part of the next quarterly report.

THE ASTER POLAR CLOUD MASK VERSION 1 ALGORITHM THEORETICAL BASIS DOCUMENT

Ronald M. Welch, Manuel A. Penaloza, and Rand E. Feind
Institute of Atmospheric Sciences
South Dakota School of Mines and Technology
501 E. St. Joseph Street
Rapid City, South Dakota 57701-3995

1.0 INTRODUCTION

One of the most important environmental challenges facing mankind is the problem of climate change. While it is recognized that mankind's activities affect the global environment, a quantitative assessment of the magnitude of these changes presently is beyond reach. A thorough description and understanding of processes at the earth's surface and in the atmosphere is necessary before realistic climate prediction can be realized. A quantitative assessment of the magnitude of potential global and regional changes must be made before policy makers are likely to be persuaded to make difficult economic decisions.

With the growing awareness and debate over the potential changes associated with global climate change, the polar regions are receiving increased attention. Since greenhouse forcings are expected to be amplified in polar regions (Wetherald and Manabe, 1986; Schlesinger and Mitchell, 1987; Steffen and Lewis, 1988), these regions may act as early warning indicators of actual climate shifts.

Global cloud distributions can be expected to be altered by increased greenhouse forcing. In the polar regions, cloud cover changes can be expected to have a significant effect on sea ice conditions (Shine and Crane, 1984) and on regional ice-albedo feedback (Barry *et al.*, 1984). In particular, polar stratus is very important to the polar heat balance and directly affects surface melting (Parkinson *et al.*, 1987). However, in order to monitor changes in polar surface conditions and polar cloudiness, more accurate procedures must be developed to distinguish between cloud and snow-covered surfaces.

Owing to the similarity of cloud and snow-ice spectral signatures in both visible and infrared wavelengths, it is difficult to distinguish clouds from surface features in the polar regions. In the visible channels, thin ice, ice fragments, wet ice, and pancake ice have low albedos and can be misinterpreted as water, melt ponds, or as thin cloud/haze. Persistent surface inversions and low clouds in winter, and near isothermal structure and extensive stratiform clouds in summer, limit discrimination in the infrared channels. Clearly, spectral signatures alone appear to be inadequate for polar scene identification (McGuffie *et al.*, 1988; Rossow *et al.*, 1989; Stowe *et al.*, 1989).

Textural signatures can be used to distinguish between cloud, snow-covered mountains, solid and broken sea ice and floes, and open water (Welch *et al.*, 1990). Likewise, textures derived from high spatial resolution imagery can be used to distinguish between various cloud types (Welch *et al.*, 1988; Kuo *et al.*, 1988). The combination of spectral and textural features has proven to be effective for polar scene classification (Ebert, 1987, 1989; Key, 1990; Welch *et al.*, 1990, 1992; Rabindra *et al.*, 1992; Tovinkere *et al.*, 1993).

Artificial intelligence (AI) increasingly is being used for classification (Key *et al.*, 1989; Lee *et al.*, 1990; Rabindra *et al.*, 1992; Tovinkere *et al.*, 1993). However there is little information as to the strengths and limitations of various AI architectures. Tovinkere *et al.* (1993) examined six different AI approaches for polar scene identification. They are (1) a feed forward back propagation neural network; (2) a probabilistic neural network; (3) a hybrid neural network; (4) a "don't care" feed forward perceptron model; (5) a "don't care" feed forward back propagation neural network; and (6) a fuzzy logic based expert system. For the present algorithm development, we will focus upon fuzzy logic and the "don't care" neural network.

The ASTER Cloud Mask Algorithm is built upon a well-established foundation of spectral and textural features which utilizes AI techniques to enhance classification accuracy. There is a widely-held notion that AI techniques require orders of magnitude greater computer resources than do the more traditional classifiers such as Maximum Likelihood. It should be noted here at the outset that this perception is totally false. While it is true that some AI techniques require a great deal more CPU time in their training phases, they require about the same number of CPU's as traditional methods in the operational environment.

This Algorithm Theoretical Basis Document is organized as follows. Section 2 includes the instrument characteristics, the labeling procedure, a description of the spectral features, texture, and several artificial intelligence approaches used for classification. Section 3 contains the algorithm description, including preprocessing, preliminary classification, fuzzy expert classification, spatial context consistency tests and quality assurance tests. Finally, Section 4 considers the constraints, limitations, and assumptions used in the algorithm.

2.0 OVERVIEW AND BACKGROUND

Cloud masking is of critical importance in the polar regions. Clouds and snow/ice surface features have similar characteristics in both the visible and infrared spectral ranges, thereby making polar scene identification extremely difficult. Often both high level cirrus clouds and low level fog and stratus clouds are optically thin. In such cases, retrievals of surface parameters such as albedo and temperatures are seriously compromised.

As part of the International Satellite Cloud Climatology Project (ISCCP), global cloud cover is being retrieved operationally. Typical cloud masking algorithms assume

that clouds can be detected using visible and infrared channel thresholds. Reflectance thresholds typically are set about 3% above the background, and thermal thresholds typically are set about 3°C below the background. Other approaches rely upon bispectral thresholding (Minnis and Harrison, 1984) and a variety of statistical methods (e.g., Saunders and Kriebel, 1988). However, Rossow *et al.* (1989), Stowe *et al.* (1989) and many other have reported difficulties associated with polar cloud cover retrievals. Indeed, LANDSAT imagery shows that clouds often are darker than the background snow and ice (Welch *et al.*, 1990). In particular, cloud cover often is confused with melt ponds, thin ice and pancake ice. In the infrared spectrum, low surface temperatures, strong inversions and isothermal structure make cloud discrimination difficult.

There are three main factors that must be addressed in the development of an operational polar cloud masking algorithm: (1) the choice of the feature vector, (2) the choice of the classifier, and (3) proper identification and labeling of regions. First, from the wide range of techniques employed by different groups, it can be seen that there is no consensus as to what type of signatures are needed in a robust polar cloud/surface feature classification scheme. However, it is generally recognized that spectral information alone is inadequate. Difference and ratio channels and textural measures need to be explored further to determine if there is an optimum feature vector for polar scene identification. In this regard, Gray Level Difference Vector textural measures seem to offer a good combination of discriminating ability and low storage and CPU requirements.

There is no consensus as to the size of the region over which textures should be computed. A new paper by Nair and Welch (1994) suggests that 16 x 16 pixel regions are preferred. Selection of regions that are too small leads to unstable textural measures. However, selection of regions which are too large leads to loss of information; i.e., the region is progressively more likely to contain multiple classes. The 16 x 16 pixel region seems to provide the best compromise between textural stability and discrimination of features. As a caveat it should be noted that textural measures based upon third (cluster shade) and fourth (cluster prominence) order statistics generally do not attain textural stability with 16 x 16 pixel templates; considerably larger template sizes are required for them. Therefore, these measures should be applied only after stability analysis is performed.

AI classifiers can significantly increase classification accuracy. The main drawbacks of the AI schemes are that these methods are not well understood by most groups and that the training time may be longer than for traditional approaches. On the other hand, in an operational mode the AI approaches are very fast. It is also worth noting that the AI methods are both nonlinear and nonparametric. Thus, no particular class structure (i.e., normal distribution) is assumed; rather, the classifier learns through presentation of examples. The AI approaches also require far fewer training samples than do traditional schemes, and the new AI approaches provide very high classification accuracies. The most impressive of these techniques to date is the "Don't Care" Back Propagation neural network (Welch *et al.*, 1992). Indeed, for regions containing pure

classes, the accuracy exceeds 95%. Nevertheless, more work is needed to extend these results to mixed classes.

A final issue concerns scene identification and labeling. No classification scheme can be expected to produce accurate results if the labeling is incorrect. The old adage "garbage in, garbage out" is especially appropriate here. The analyst needs to examine a wide variety of information before labeling a region. A new Interactive Visual Image Classification System (IVICS) has been introduced which provides a wide variety of analysis tools to the user. This system has greatly facilitated the selection of pure training samples and accurate labeling. This is particularly important in polar scene analysis where erroneous labeling is problematic. By virtue of using this system, very high classification accuracies (>95%) been attained for polar regions.

This section includes:

- Experimental Objectives and Data. The basic elements of the cloud mask algorithm are described here, along with the data and form of the output.
- Historical Perspective
- Instrument Characteristics
- Pixel Labeling
- Spectral Features
- Texture
- Artificial Intelligence Classifiers

2.1 Experimental Objective and Data

This is Version 1 of the ASTER Polar Cloud Mask. Polar regions are defined in this algorithm to consist of all land and water regions lying poleward of 60°N or 60°S. Version 1 is designed to use LANDSAT, TM, AVIRIS, TIMS, and MAS data. The algorithm will be tested on all of the polar data that we can acquire, including at least 24 LANDSAT TM scenes over Antarctica and all of the MAS scenes acquired by Mike King over the Beaufort Sea. It will be modified as needed for Release 2 and delivered for use with ASTER data.

The ASTER Polar Cloud Mask in part is designed as a validation tool for MODIS. Several members of the ASTER cloud mask development team also are involved with the CERES and MODIS global cloud mask algorithm development. Therefore, close coordination between the ASTER, CERES, and MODIS efforts will be maintained.

The ASTER polar cloud masking algorithm relies heavily upon a heritage of experience with AVHRR, LANDSAT, and AVIRIS data analysis. The algorithm consists of the following elements (Fig. 1):

1) Preprocessing. This step locates missing lines and stripes in the data and then performs reconstruction. This step also normalizes the reflectances for solar zenith angle.

2) Table Lookup Classification. This is a fast, preliminary classification of unambiguous feature vectors located close to cluster centers. Only spectral features, including band ratios and band differences, are used in this step.

3) Fuzzy Expert Classification. This step utilizes combinations of spectral and textural features in a fuzzy logic artificial intelligence classifier. It also makes use of the following ancillary data:

- 500 m resolution coastline data, with lakes and rivers.
- 10 minute resolution topographical data.
- 10 minute resolution ecosystem map.
- 18 km resolution U.S. Navy/NOAA weekly Sea Ice Product.
- 150 km resolution weekly NOAA Snow Data Product.
- 1 km resolution MODIS daily snow and sea ice mask (available after launch).
- NMC Surface Temperatures

4) Spatial Consistency Test. In this step the classification of each pixel is compared to that of neighboring pixels and with the databases (ecosystem, type, elevation, snow-cover).

5) Quality Assurance Test. A small fraction of the scene (<1%) will be randomly selected for reclassification with a hierarchical don't care neural network. Statistics are accumulated, and the quality of the classification is assessed, based upon the statistics. If necessary, the scene is flagged for human expert evaluation; otherwise, a 1-byte cloud mask is produced.

For Version 1 each pixel will be classified using a bit map:

unknown	thick cloud	thin cloud	shadow	ice	wet ice	water	land

All bits are set to zero, except as appropriate. For example, for a case of thin cloud over broken sea ice, the thin cloud, ice and water bits would be set to a value of unity, with all other bits set to zero. At present, the wet ice bit includes melt ponds, slush ice, and thin ice. These features have values of reflectivity intermediate between that of water and ice/snow, and with temperatures near 0°C. It is possible to identify specific features such

ALGORITHM FLOW

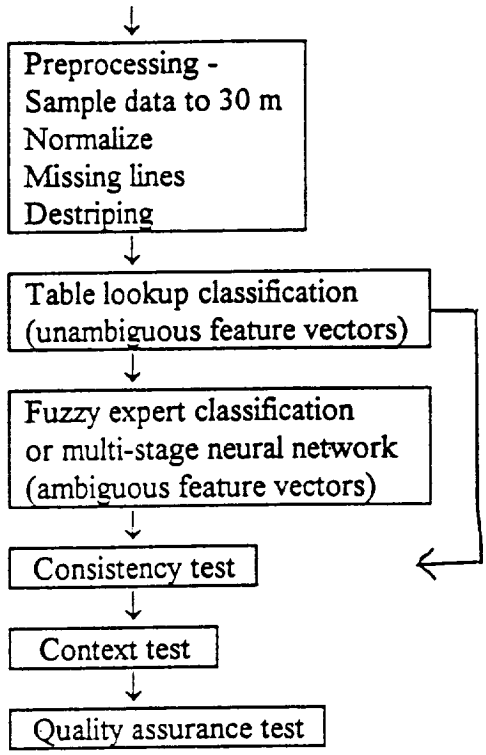


Fig. 1: Algorithm flow chart

as melt ponds, leads, and slush ice, but this is not planned in the current algorithm development. It is also possible to distinguish between water and ice cloud, but this also is not included here.

Often pixels will be comprised of mixed, classes, and it is potentially useful to indicate the fractional presence of each class. An alternative bit map that we propose for Version 2 enables recording of class percentages and reserves 2 bits per class. These 2 bits would code percentage ranges of a class within a pixel. The following table illustrates the use of the 2 bits.

bits	description
00	class present with the range 0-10%
01	class present within the range 10.01-50%
10	class present within the range 50.01-90%
11	class present within the range 90.01-100%

This scheme doubles the size of the bit map but it provides additional information to researchers processing scenes with specific requirements. For example, an investigator interested in surface samples at a high degree of confidence would only select bit combinations of 11. However, another investigator interested in cloud properties might select all pixels with cloud present at in fractions greater than 50% (i.e., 10 and 11 bit combinations).

Much of our early algorithm development has been based upon AVHRR LAC data. However, the current algorithms are based upon 24 LANDSAT TM Quarter scenes. MODIS Airborne Sensor data acquired over the Beaufort Sea has been requested from Mike King and also will be used in this algorithm development.

Figures 2 and 3 each show nine of the LANDSAT TM scenes that we are using. These scenes are all taken in or near Antarctica. These scenes are displayed as histogram equalized three-band overlays, with channel 6 (infrared) as red, channel 5 as green and channel 4 as blue. Note that a wide range of conditions are included: thin cirrus, thin and thick stratocumulus, cumulus, ice-covered land, broken sea ice, slush, and melt ponds. To ensure that the cloud mask algorithm is robust, we need a series of LANDSAT TM scenes for the Arctic, including land areas. We also expect to obtain some MAS data of the Beaufort Sea during the first half of 1994 which will be used for the same purpose.

We are using LANDSAT TM Bands 2-7, as they correspond closely to the following planned ASTER bands:

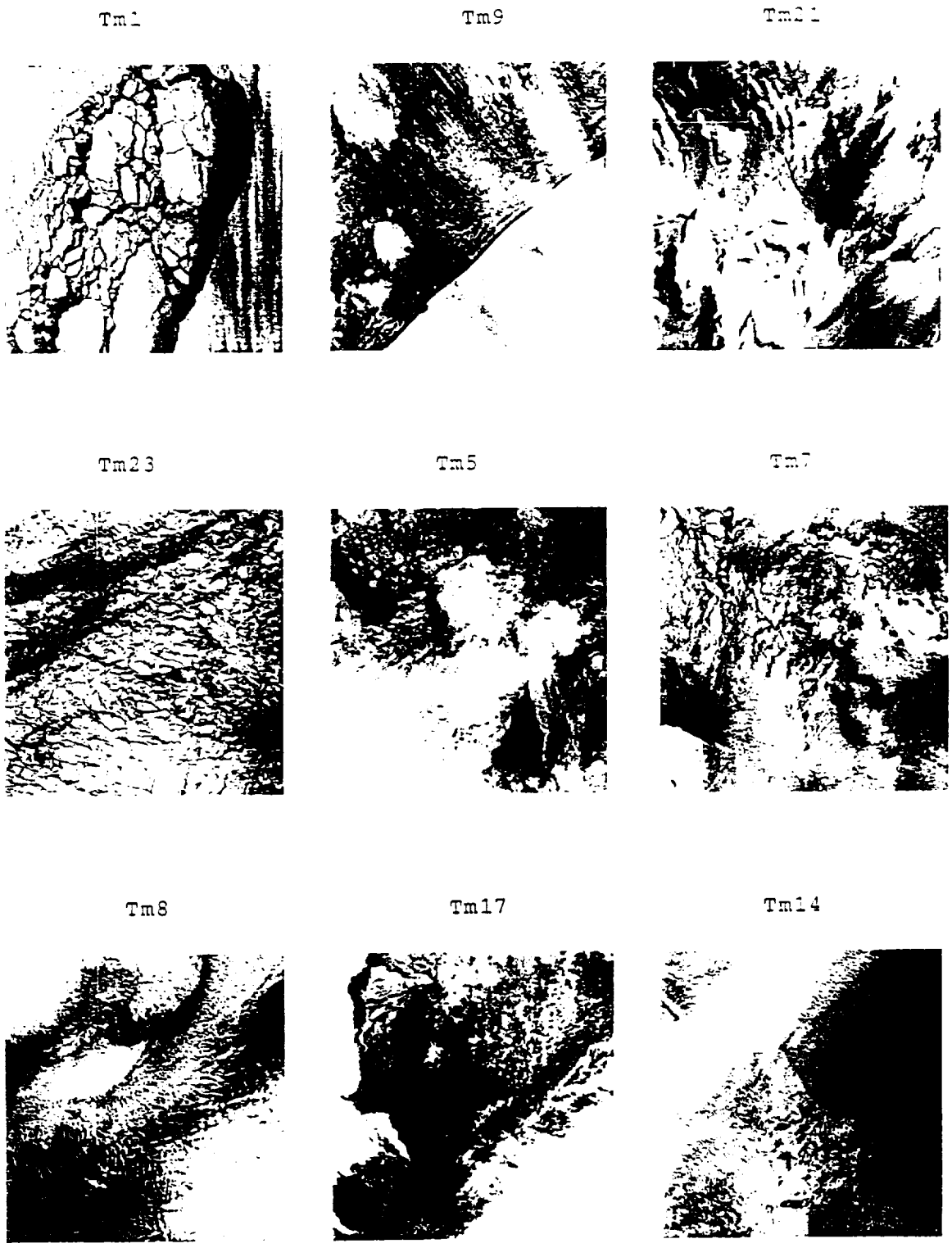


Figure 2

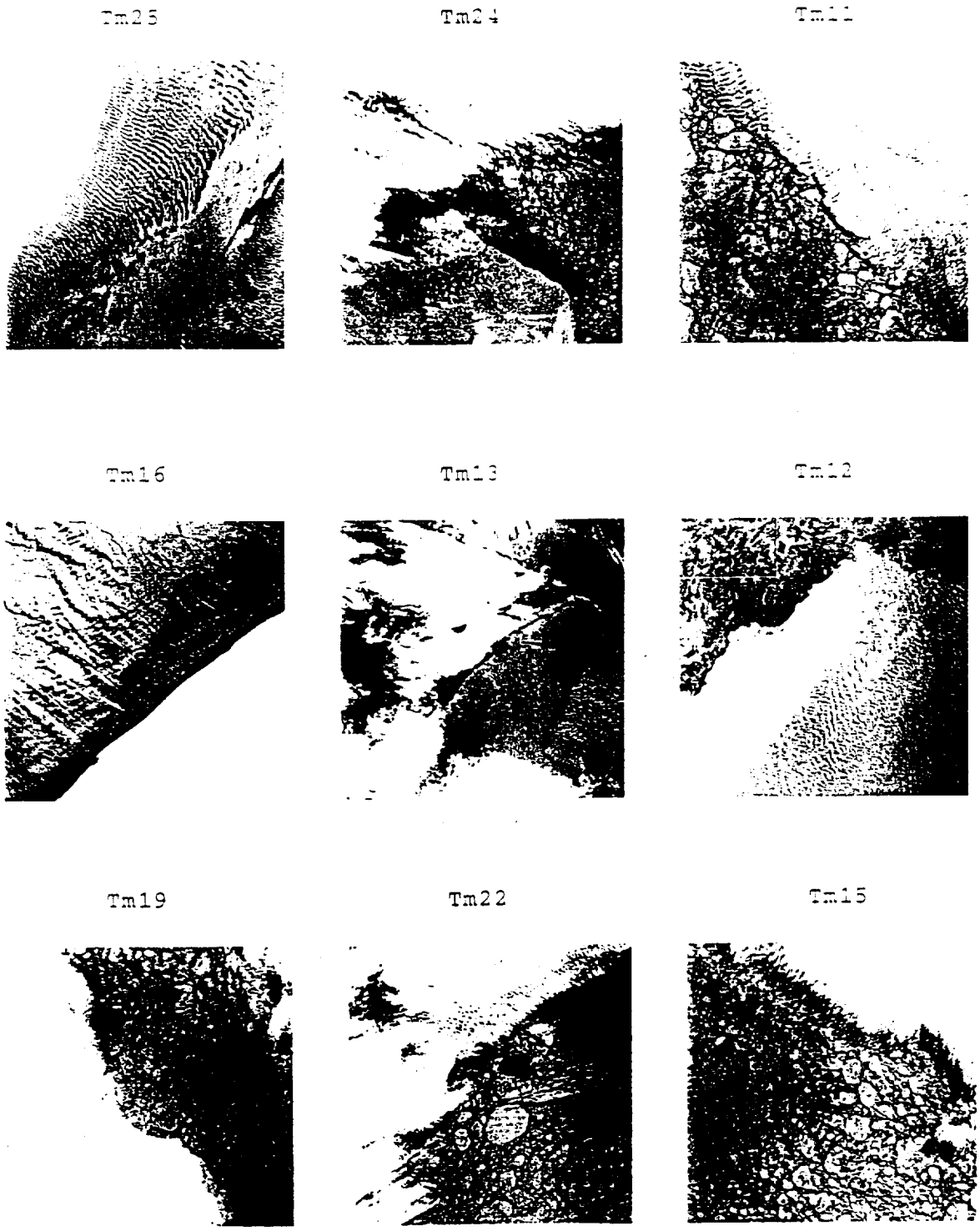


Figure 3

TM	ASTER
2 (0.52-0.60 μm)	1 (0.52-0.60 μm : VNIR)
3 (0.63-0.69 μm)	2 (0.63-0.69 μm : VNIR)
4 (0.76-0.90 μm)	3 (0.76-0.86 μm : VNIR)
5 (1.55-1.75 μm)	4 (1.60-1.70 μm : SWIR)
6 (10.4-12.5 μm)	13 and 14 (10.25-10.95 μm and 10.95-11.65 μm : TIR)
7 (2.08-2.35 μm)	5-8 (2.145-2.185 μm, 2.185-2.225 μm, 2.235-2.285 μm, and 2.295-2.365 μm : SWIR)

The final ASTER algorithm probably will use only channels (i.e., 1, 3, 4, 5, 10, 13, 14). The other channels will be used as substitutes when either the primary choices are bad or for redundancy tests.

2.2 Historical Perspective

In this section, a number of key aspects to this algorithm development and in-house experience with them are summarized. Each one provides a piece of the classification strategy for the algorithm. However, the experience of many other investigators is also incorporated in the classification methodology described in Section 3.0. Some of them include: Key, 1989; Li and Leighton, 1991; Ebert, 1987, 1989, 1992; Key *et al.*, 1990; Allen *et al.*, 1989; Saunders and Kriebel, 1988; Raschke *et al.*, 1992; McGuffie *et al.*, 1988; Ormsby and Hall, 1991; Sakellariou and Leighton, 1988; Crane and Anderson, 1984; Simpson and Humphrey, 1990; King and Tsay, 1993; Menzel and Strabala, 1993; Welch *et al.*, 1988, 1990, 1992; Tovinkere *et al.*, 1993; Rossow, 1989, 1993; Stowe *et al.*, 1991; Rossow *et al.*, 1989; Seze and Rossow, 1991; Kuo *et al.*, 1988.

Much experience is based on low resolution (1-4 km) AVHRR non-polar imagery. Relatively little work has been performed on high spatial resolution polar imagery. In particular, there is virtually no experience with high spatial resolution polar imagery for the full compliment of ASTER channels (especially multispectral imagery in the thermal IR region of the spectrum from 8-12 μm). Some classification methodologies do not require features from thermal IR region and are only dependent on solar spectral region features; therefore, methodologies applied to low spatial resolution polar imagery in the past are certainly germane here. We expect the most significant differences will occur for textural features. Higher spatial resolution imagery can manifest higher spatial contrast and frequencies. As we gain additional experience with high resolution imagery, we expect that the sets of textural features proposed herein and the parameters used to compute them will be modified somewhat.

2.3 Instrument Characteristics

ASTER will provide data in the three spectral regions using three separate radiometer subsystems. These are the visible and near-infrared (VNIR) subsystem being provided by NEC, the short wavelength infrared (SWIR) subsystem provided by MELCO,

and the thermal infrared (TIR) subsystem provided by FUJITSU. The instrument band passes, radiometric accuracies, and radiometric and spatial resolution are given in Table 1. The VNIR includes a separate, single-spectral-band (0.76-0.86 μm, channel 3B) radiometer inclined backward at an angle of 27.7° to the other sensors to provide a 15-m same-orbit stereoscopic imaging capability. A wide dynamic range and multiple gain settings will help ensure useful data for a variety of investigations.

The swath width for all three systems is 60 km. The ASTER instrument has a crosstrack pointing capability of 8.55° for the SWIR and TIR subsystems, and 24° for the VNIR subsystem. This gives crosstrack observing ranges on the ground of approximately ≈136 km and ≈343 km respectively, ensuring that any point on the globe will be accessible at least once every 16 days for the SWIR and TIR, and once every five days for the VNIR. However, in most instances, all three radiometer systems will image the same 60-km ground swath.

Table I. Spectral and spatial characteristics of the Advanced Spaceborne Thermal Emission Reflectance Radiometer (ASTER). Asterisk indicates the stereo band. Stereo B/H ratio 0.6

ASTER					
Wavelength Region	Band Number	Spectral Range	Radiometric Accuracy	Radiometric Resolution	Spatial Resolution
VNIR	1	0.52-0.60	+/- 4%	≤0.5%	15m
	2	0.63-0.69	+/- 4%	≤0.5%	15m
	3	0.76-0.85*	+/- 4%	≤0.5%	15m
SWIR	4	1.60-1.70	+/- 4%	≤0.5%	30m
	5	2.145-2.185	+/- 4%	≤1.3%	30m
	6	2.185-2.225	+/- 4%	≤1.3%	30m
	7	2.235-2.285	+/- 4%	≤1.3%	30m
	8	2.295-2.365	+/- 4%	≤1.0%	30m
	9	2.360-2.430	+/- 4%	≤1.3%	30m
TIR	10	8.125-8.475	1-3K	≤0.3K	90m
	11	8.475-8.825	1.3K	≤0.3K	90m
	12	8.925-9.275	1-3K	≤0.3K	90m
	13	10.25-10.95	1-3K	≤0.3K	90m
	14	10.95-11.65	1-3K	≤0.3K	90m

2.4 Pixel Labeling

A critical aspect of the algorithm development is pixel and subregion labeling. Accurate labeling is the key to accurate classification. Therefore, it is important to provide the analyst with as much information as possible. Figure 4 shows an example of

tm3 r:6 g:5 b:4

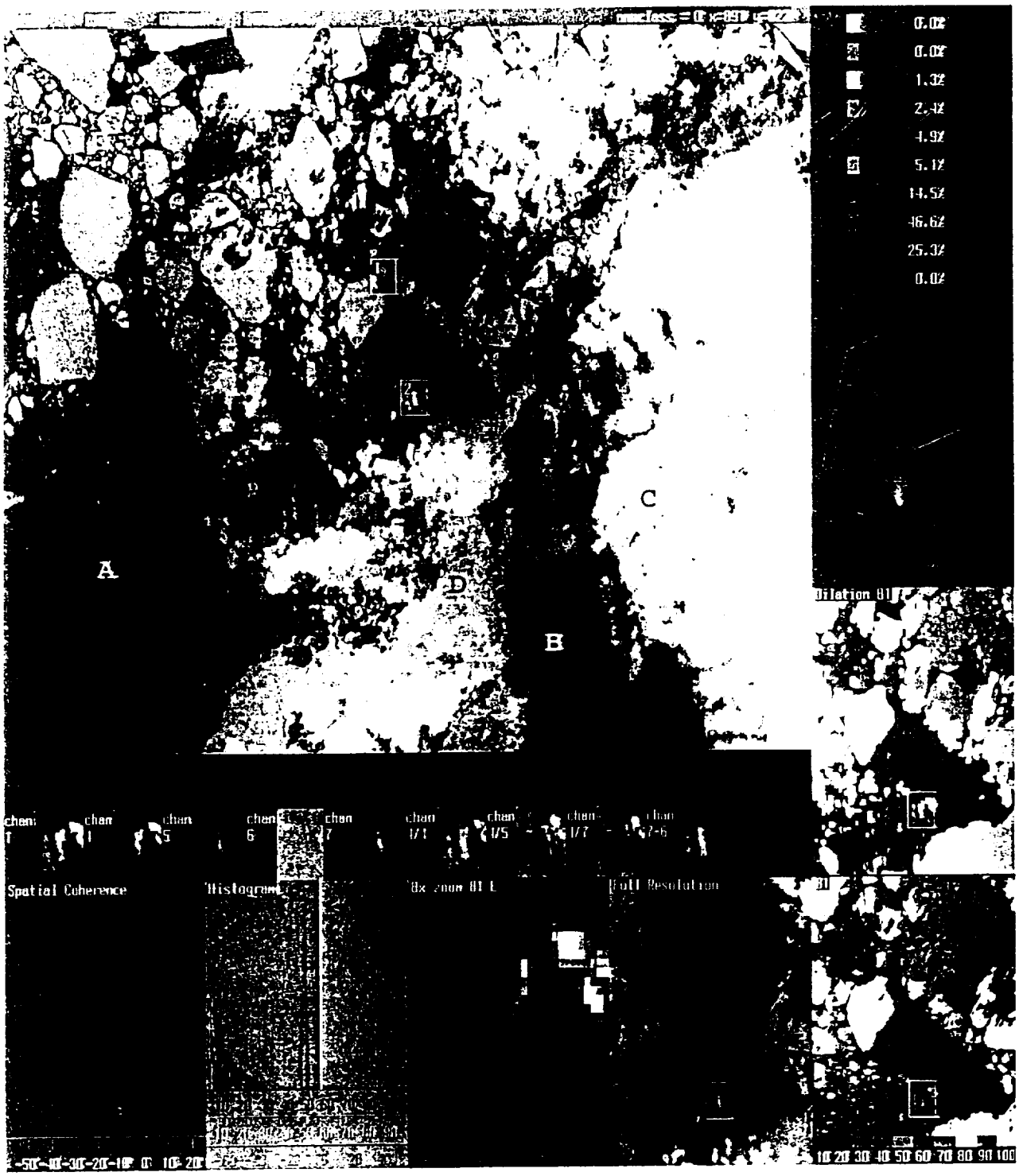


Figure 4

the Interactive Visual Image Classification System (IVICS) which displays three-band color overlays. A series of pull-down menus are available to the analyst which allow a wide range of channel displays and image processing functions. By default all bands are histogram equalized for contrast enhancement. However, any combination of band differences and band ratios can be designed and displayed on command. Additional display features such as principal components, decorrelation stretch, canonical transformations, and edge finding are being implemented.

The large central image is a full spatial resolution subsample of the original image. The region labeled "A" is water, "B" is shadow on ice, "C" is stratocumulus cloud, "D" is ice-covered land, and "E" is broken sea ice. Directly under the central image are ten small regions which display channels 1, 4, 5, 6, 7, 4/1, 4/5, 4/7, and 7-6. The analyst immediately can examine the region outlined in the box in each of these channels. These 10 small regions also are used as icons for mouse control, as explained below. Starting at the lower left corner of the monitor and moving to the right, a series of special purpose displays are provided. First a spatial coherence (2 x 2 pixel plot of mean versus standard deviation) (Coakley and Bretherton, 1982) is provided. Next are histograms of all three channels (shown in red, green, and blue), where channels 4 (blue) and 5 (green) are displayed in terms of albedo, and channel 6 (red) is displayed in terms of temperature (°C). In the lower center is a 8x zoom of channel 1, followed by a three-color enlarged display of the selected regions. At the bottom right, a color density sliced version of channel 1 is shown; the percentage of pixels in each color range is given at the top right. In addition, to the far right, a morphological dilation (Serra, 1982) is shown for channel 1. The analyst has various options for each of these special displays which are activated by clicking the mouse on any of the ten icons. Finally, three-dimensional cluster analysis is shown in the center right cube; this cube displays 3-D clusters, computed from the selected region (i.e., box), for the red, green, and blue bands. The cluster cube rotates continuously, but can be halted with the mouse button. A trained analyst can see immediately from the three-dimensional cluster display if the boxed region contains pure classes. In Fig. 3 the boxed region labeled as "2" in the main display contains both cloud and water, which is clearly differentiated in the 3-D cluster display.

2.5 Spectral Features

Clouds generally are characterized by higher albedos and lower temperatures than the underlying surface. However, there are numerous conditions when this characterization is inappropriate, most notably over snow and ice. Of the cloud types, cirrus, low stratus and small cumulus are the most difficult to detect. Likewise, cloud edges are difficult to recognize when they do not completely fill the instantaneous-field-of-view (IFOV), i.e., pixel.

The NOAA Cloud AVHRR (CLAVR) algorithm uses all five channels of AVHRR to derive a global cloud mask (Stowe *et al.*, 1989). It examines multispectral information, channel differences, and spatial differences and then employs a series of sequential decision tree tests. Cloudfree, mixed (variable cloudy) and cloudy regions are identified

for 2 x 2 GAC pixel arrays. If all four pixels in the array fail all the cloud tests, then the array is labeled as cloudfree (0% cloudy); if all pixels satisfy all the cloud tests, then the array is labeled as 100% cloudy. If 1 to 3 pixels fail the cloud tests, but are restored to clear by restoral tests, then the array is labeled as mixed; if the restoral tests fail, then the array is labeled as variable and assigned an arbitrary value of 50% cloudy. The set of cloud tests is subdivided into Daytime Ocean Scenes, Daytime Land Scenes, Nighttime Ocean Scenes, and Nighttime Land Scenes.

The International Satellite Cloud Climatology Project (ISCCP) cloud masking algorithm is described by Rossow (1989, 1993), Rossow *et al.* (1989) and Seze and Rossow (1991). Only two channels are used, the narrowband VIS (0.6 μm) and the IR (11 μm). Each observed radiance value is compared against its corresponding Clear-Sky Composite value. This step uses VIS radiances, not VIS reflectances. Clouds are assumed to be detected only when they alter the radiances by more than the uncertainty in the clear values. In this way the "threshold" for cloud detection is the magnitude of the uncertainty in the clear radiance estimates. As such this algorithm is not a constant threshold method such as used in the CLAVR algorithm.

The ISCCP algorithm is "based on the premise that the observed VIS and IR radiances are caused by only two types of conditions, 'cloudy' and 'clear', and that the ranges of radiances and their variability that are associated with these two conditions do not overlap" (Rossow, 1993). As a result, the algorithm is based upon thresholds, where a pixel is classified as "cloudy" only if at least one radiance value is distinct from the inferred "clear" value by an amount larger than the uncertainty in that "clear" value. The uncertainty can be caused both by measurement errors and by natural variability. This algorithm is constructed to be "cloud-conservative", minimizing false cloud detections but missing clouds that resemble clear conditions.

The present algorithm borrows from the CLAVR and ISCCP approaches, and from other sources, as noted. The following figures show preliminary results for the following ten classes:

1. Water
2. Slush/wet ice
3. Sea ice
4. Snow over land
5. Thin (semi-transparent) water cloud
6. Thick (opaque) water cloud
7. Thin cirrus
8. Thick cirrus
9. Land
10. Shadows over ice

2.5.1 Daytime Scenes

Figure 5 shows scatter diagrams for several different band combinations: band 5/band 2 versus band 3, band 6 versus band 4-band 5, and band 7/band 4 versus band 6. The upper plots show the range of scatter; the corresponding lower plots show the cluster centers within one standard deviation of the mean. The second stage of the current algorithm uses navigation to determine if the pixel is over land or ocean. Then it uses a lookup table to make a preliminary classification. The lookup tables are based upon the results shown in this section.

At present we use a very conservative 1σ standard for class differentiation, as shown in Fig. 5. This is to assure accuracy of the results. However, only about 10% of the pixels are classified with this preliminary prcedure. Further analysis is being performed to determine if the 1σ standard can be relaxed to include a larger number of pixels without severely compromising accuracy.

Figure 5 shows that with the proper selection of channels, it is possible to discriminate between water, shadows, sea ice, and thin cirrus. Note, however, that water cannot have a temperature below 271°K. The low temperature values of channel 6 are due to a LANDSAT calibration problem. To correct for this situation, the channel 6 temperatures are adjusted so that water pixels have a value of 271°K.

Figure 6 shows a second scene with the same band combinations. Note that the water pixels are about 8°K warmer than those in Fig. 5. Therefore, a different adjustment for channel 6 temperatures is required. This shows that the channel 6 sensor is not stable, causing considerable difficulties for us in developing the algorithm. Therefore, we have had to rely more heavily upon AVHRR results for the thermal channel portion of the algorithm.

Figure 7 shows a different set of channel combinations for two additional scenes. Once again, the channel 6 values for water show variability. For Figs. 5-7, channel 6 temperatures average 10-12°K below 271°K. Note that the various channel combinations permit a preliminary separation of classes. As shown in Fig. 4, the cloud shadows over ice (labeled B) are clearly distinguishable. While not shown in these figures, detection of cloud shadows over water is possible using histogram equalized band 1 data.

For the polar regions solar zenith angles range from 60° to 85°. The observation angles vary only slightly across the swath since ASTER and LANDSAT are primarily nadir-viewing instruments. Figure 8 shows semi-infinite direct-beam albedo versus wavelength for several values of direct-beam zenith angle (Wiscombe and Warren, 1981). This suggests that solar zenith angle effects should not be large for the ASTER (or LANDSAT) VIS channels. However, these effects become important for LANDSAT channels 5 and 7 (ASTER channels 4-8).

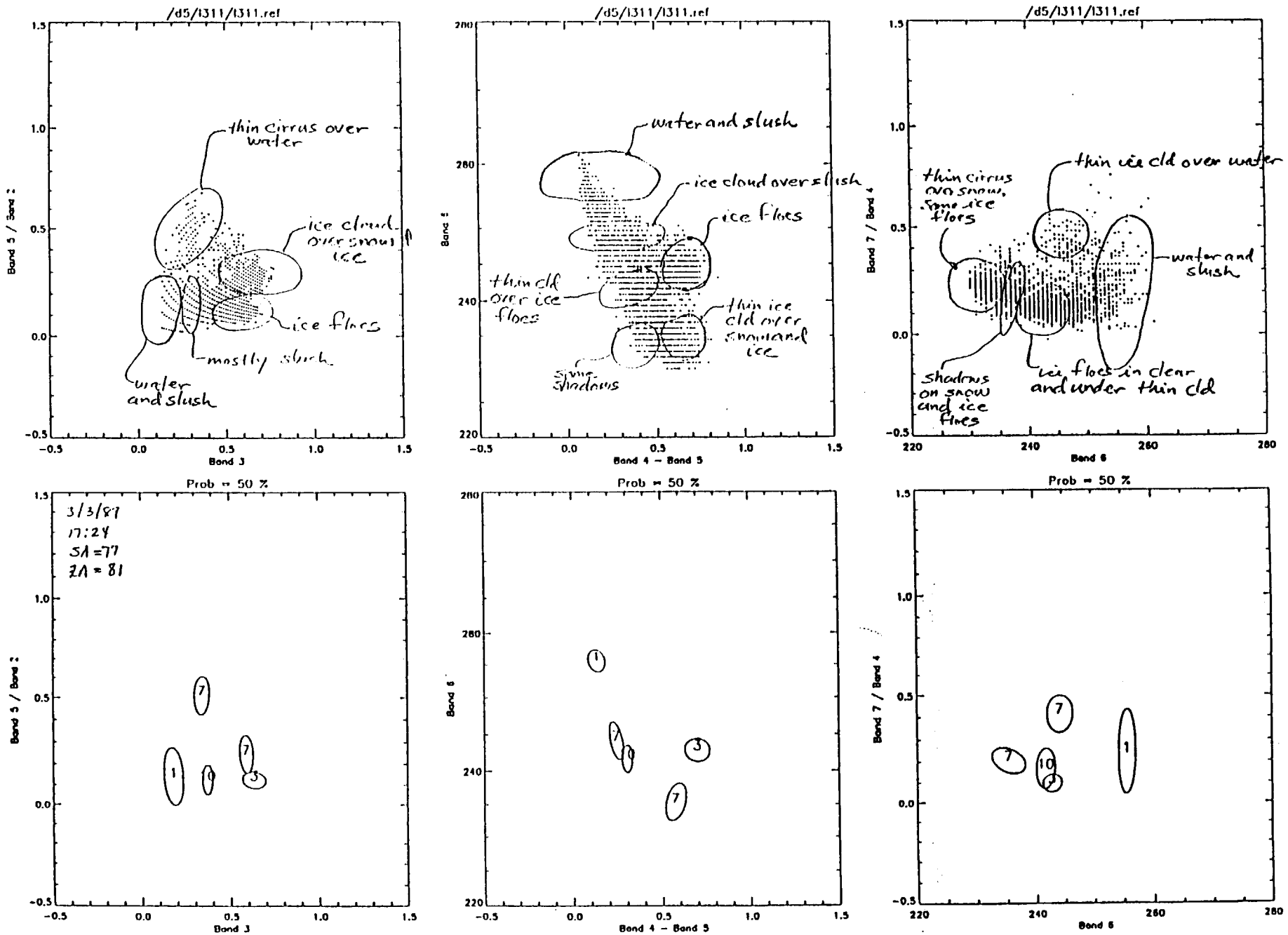


Figure 5 **Landsat TM scene L311**
Top row - Scatter for entire subsampled image
Bottom row - Training samples

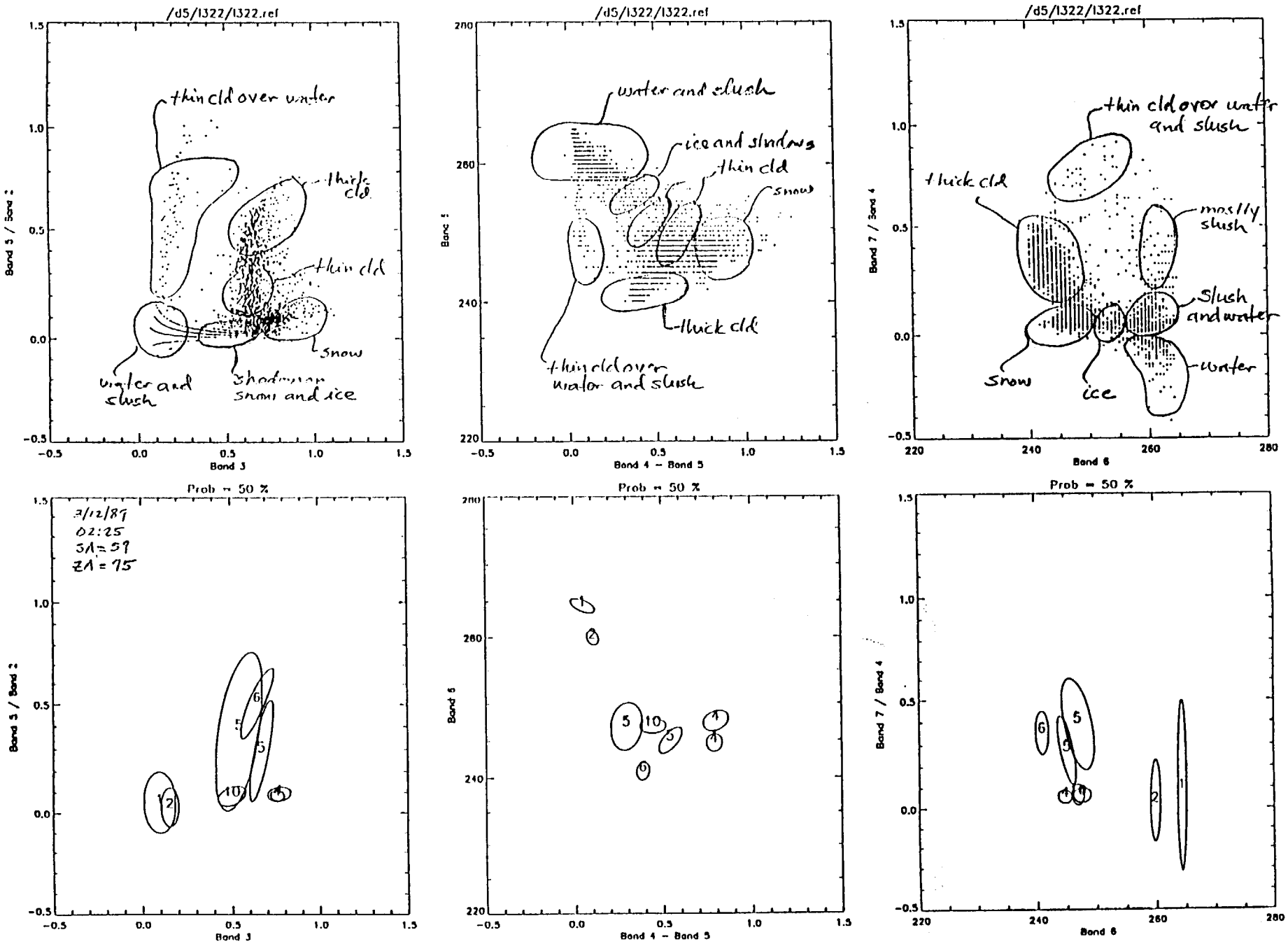


Figure 6 **Same as Figure 1 for Landsat TM scene L322**

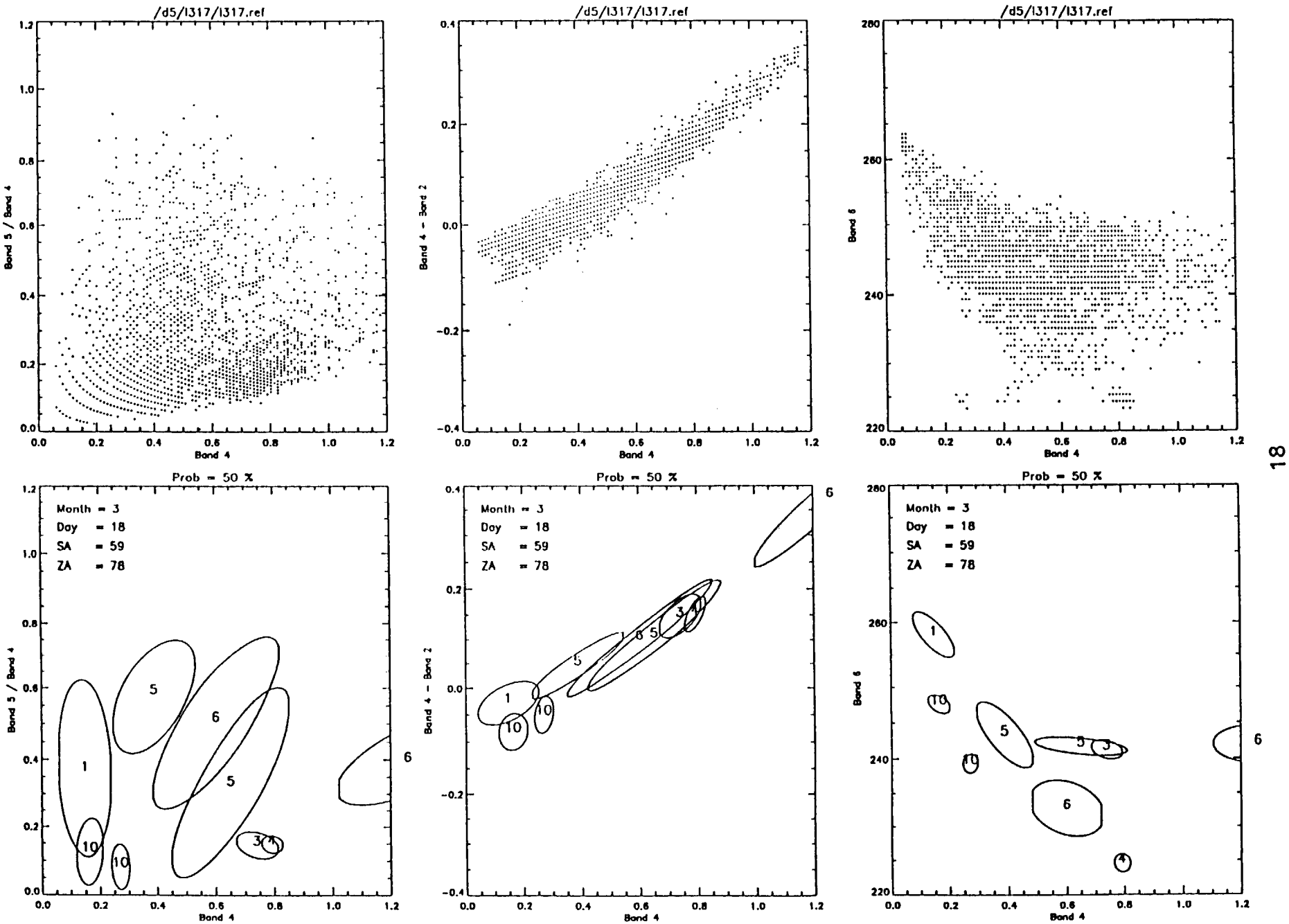


Figure 7: Same as Figure 5 but with a different set of channel combinations.

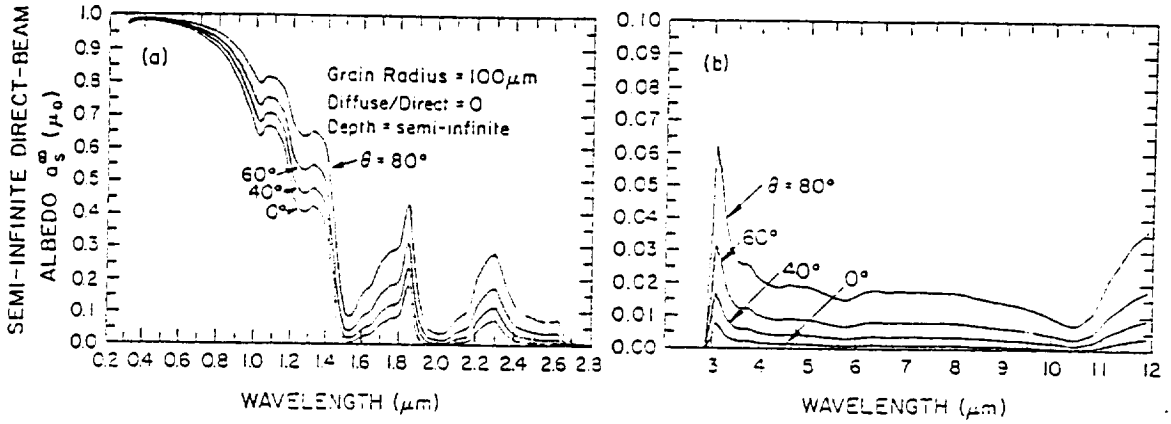


FIG. 8. Semi-infinite direct-beam albedo $a_s^*(\mu_0)$ versus wavelength for several values of direct-beam zenith angle $\theta = \cos^{-1}\mu_0$.

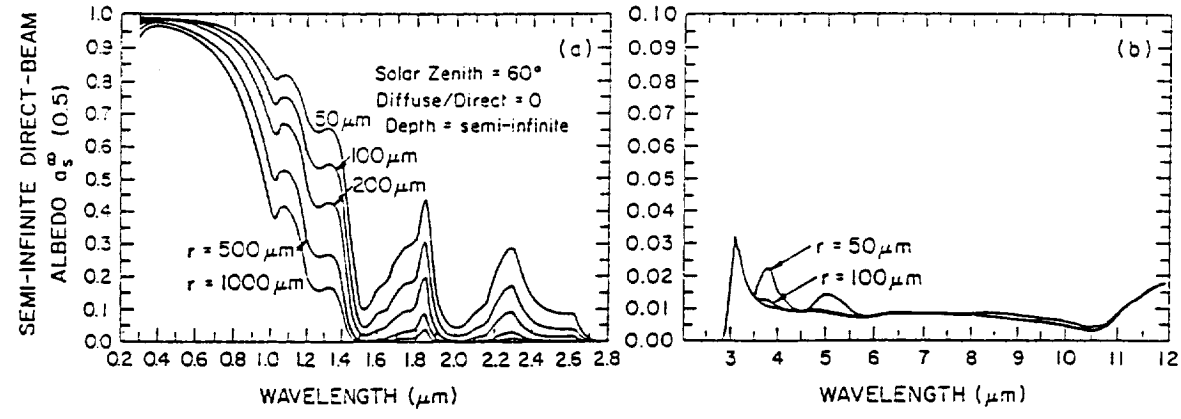


FIG. 9. Semi-infinite direct beam albedo $a_s^*(0.5)$ as a function of wavelength for various grain radii.

Figure 9 shows semi-infinite direct beam albedo as a function of wavelength for various ice grain radii. A comparison of Figs. 8 and 9 suggests that grain size (which increases with aging) causes larger albedo variations than do solar zenith angle changes from scene to scene. Therefore, the cluster centers shown in Figs. 5-7 can be expected to vary due both to solar zenith angle and ice grain size variations.

2.5.2 Nighttime Scenes

The cloud emissivity versus cloud geometric thickness is shown in Figure 10 for AVHRR channels 3-5 as a function of particle size (Yamanouchi and Kawaguchi, 1992). Note that channel 3 emissivities vary significantly from those in channels 4 and 5. The difference in emissivity is largest for smaller particles, for both water and ice clouds.

Figure 11 shows a scatter diagram of AVHRR T_4-T_5 versus T_3-T_4 (Yamanouchi and Kawaguchi, 1992), when $|T_4-T_5| \geq \Delta T_D$, then cloud is present (Yamanouchi and Kawaguchi, 1992; Stowe *et al.*, 1991). The points scatter widely for clouds, while clear pixels concentrate in a narrow region. The AVHRR T_4-T_5 test will be replaced by ASTER T_{13}-T_{14}.

ASTER does not have a 3.7 μm channel. However, it does have a 8.5 μm channel (channel 10). Figure 12, shows the complex indices of refraction of ice and water across the atmospheric window region (Warren, 1984). Minimum values in the imaginary component are found in the 8-10 μm region, indicating weak absorption. At larger wavelengths both ice and water absorb more strongly; however, ice absorbs much more strongly than water at wavelengths of 11-12 μm. Differences in the cloud radiative properties between 8.5 and 11 μm allow us to substitute the ASTER 8.5 μm channel for the AVHRR 3.7 μm channel.

Figure 13 shows that the combination of 8-11 μm versus 11-12 μm channels provides the means to distinguish between thick and thin cirrus, water cloud, and multi-layered cloud (Menzel, 1994).

Textural features are important for the nighttime cloud mask. Indeed, Welch *et al.* (1988) showed that high spatial resolution textural measures alone are sufficient to distinguish between cirrus, stratocumulus, and cumulus clouds. Textural features alone also can distinguish between cloud and surface features in the polar regions (Welch *et al.*, 1990). However, more work must be done on cloud masking using the infrared channels. Figure 14 shows AVHRR channel 4 (i.e., 11 μm) means and standard deviations for 18 polar classes (Ebert, 1987). Shown are the mean and the maximum entropy measures. The combination of spectral and textural measures is effective for polar scenes.

2.5.3 Spatial Consistency Tests

The first test, described in Rossow and Garder (1993), is similar to that of spatial coherence in that it is applied only to IR brightness temperatures. It is based upon the fact

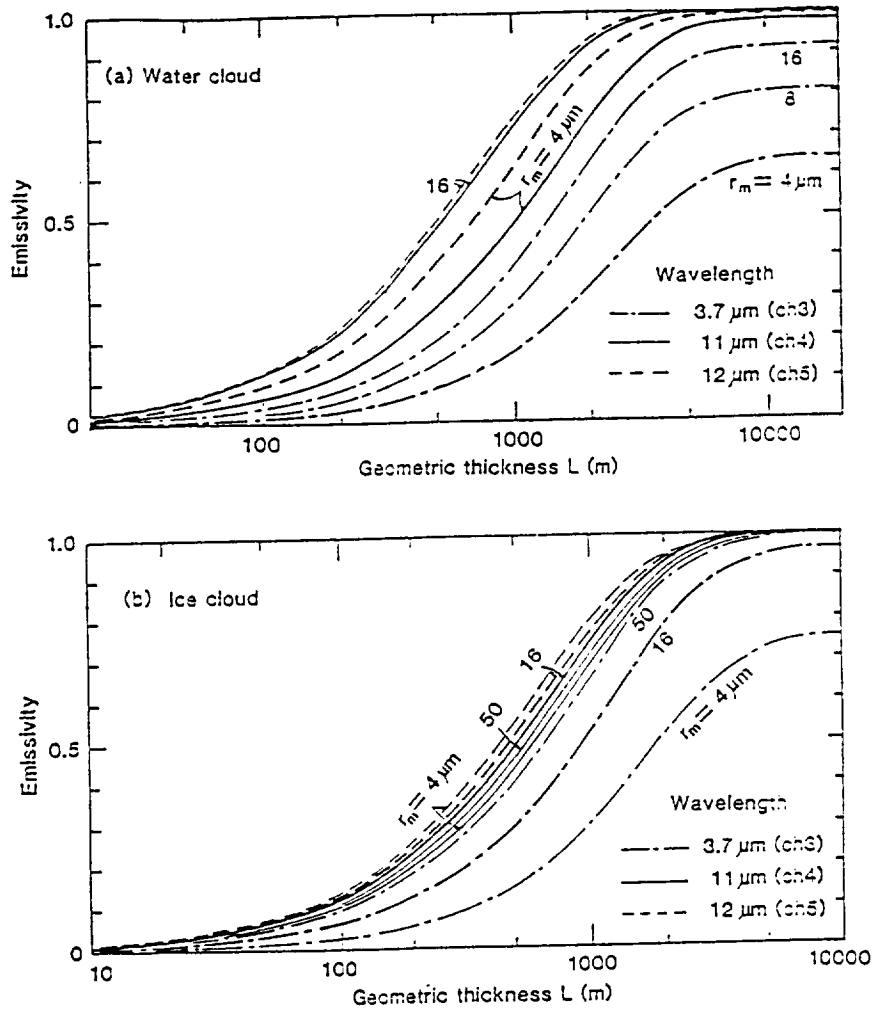


Figure 10. Emissivities of Channels 3, 4 and 5 calculated for (*a*) water droplets, and (*b*) ice spheres of several particle mode radius r_m. The abscissa is the thickness corresponding to the number density $N = 10^3/(\pi r_m^2)\,\mathrm{cm}^{-3}$, and is only arbitrary.

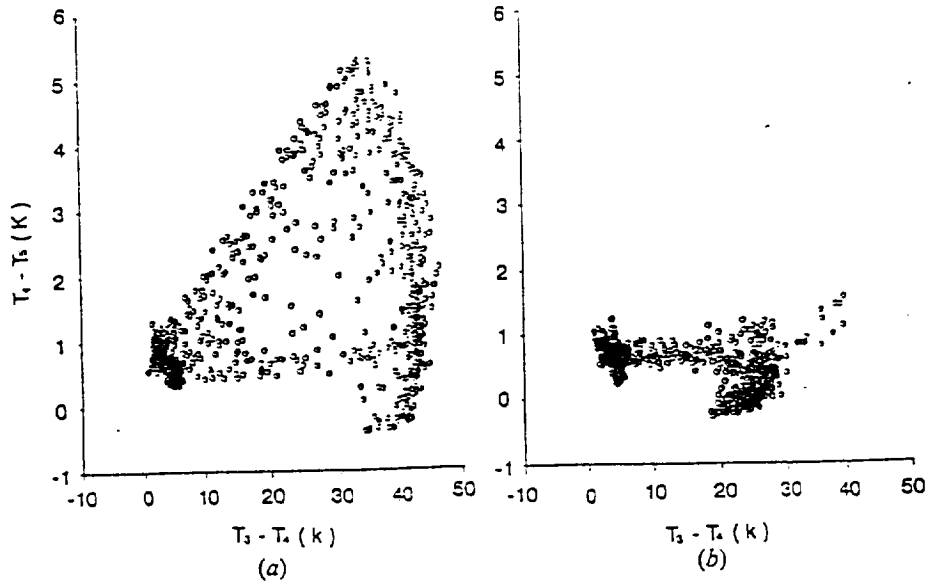


Figure 11. Scatter diagram of brightness temperature difference of channels 3 and 4 against that of channels 4 and 5 for 512 by 512 pixels area (area C) averaged for 16 by 16. (*a*) 15 December for middle-level cloud, and (*b*) 8 December for low-level cloud.

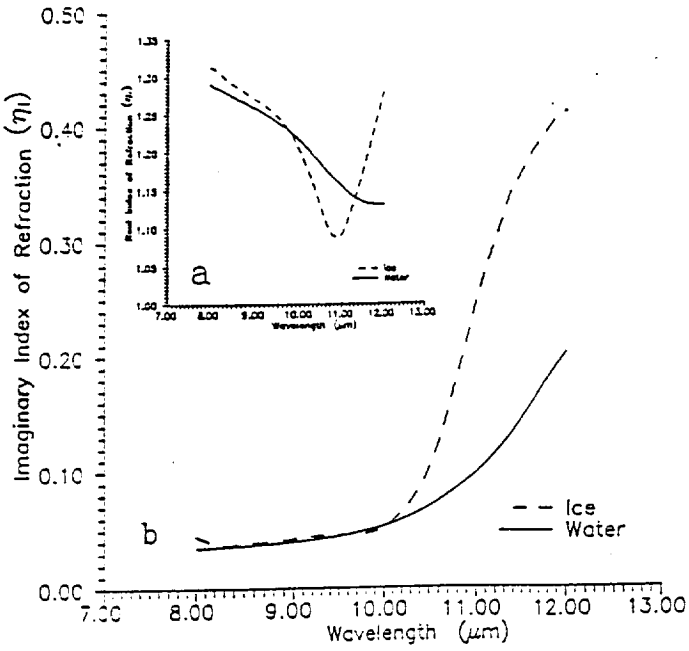


Figure 12: Indices of refraction of ice and water across the atmospheric window, (a) real portion, (b) imaginary portion.

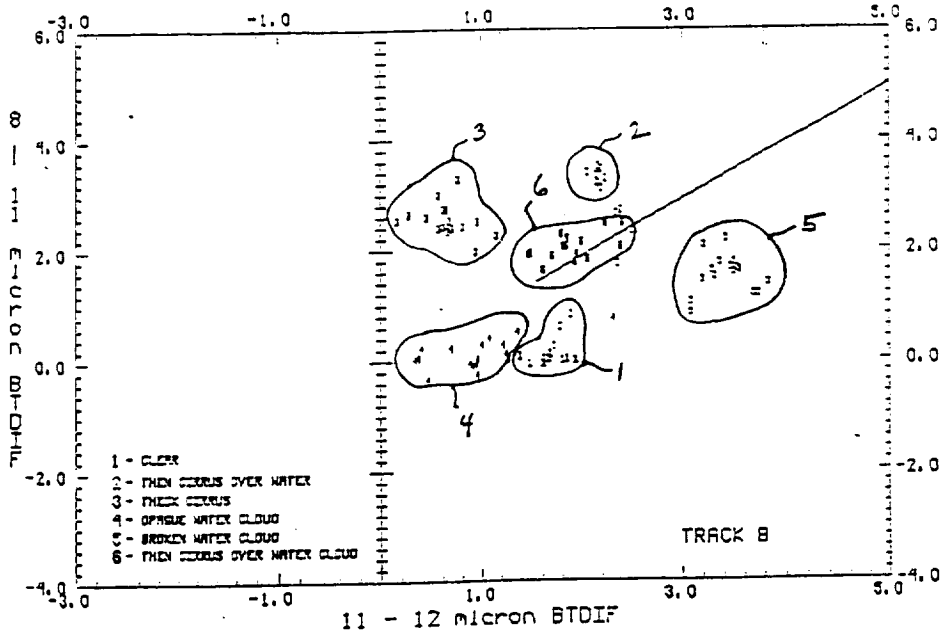


Figure 13. Same as Figure 4 except limited to the cloud scenes identified in Figure 5. Also shown for reference is a line with a slope of one.

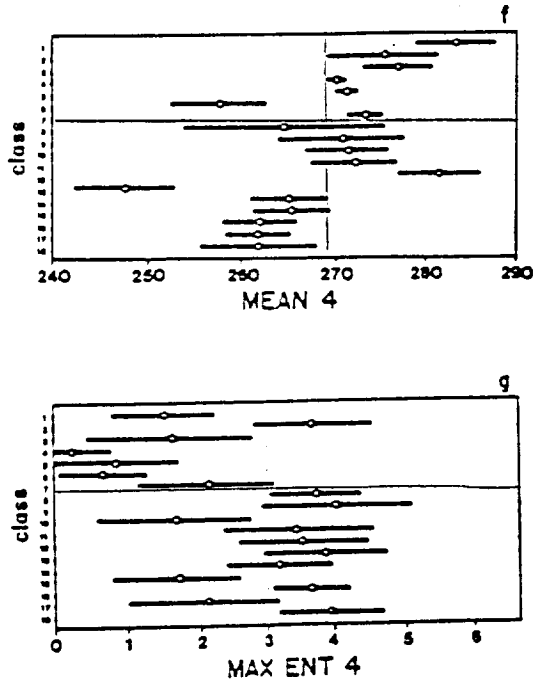


Figure 14

that clear pixels tend to be warmer than cloudy pixels and to exhibit less spatial variability. First a small local region is defined, composed of pixels with the same ecosystem. The spatial domain is approximately $(30\ km)^2$. The pixel in the local region with the largest IR (11 μm) temperature (T_{max}) is found. All pixels colder than the spatial contrast (SC) value

$$T < T_{max} - \Delta_{SC}$$

are labeled as cloudy; all others, including the warmest pixel, are labeled as "undecided". Note that not only is it important that the class pixels be identical (land or ocean), but also that the size of the region be chosen carefully. All coastal regions and all land regions containing mixed land and water pixels are excluded from this test since the inherent contrast between land and water surface radiances would dominate the results. For regions that are too large, there is increased likelihood of spatial variations in surface parameters. The shape of the test regions also can be important, since meridional gradients in surface temperature generally are larger than zonal gradients. The size of the contrast threshold must be larger than the magnitude of natural variations at the surface and smaller than that caused by clouds. Since cloud variability can be as small as surface variability, values of $\Delta_{SC} = 3.5K$ are chosen over ocean (Rossow *et al.*, 1993) and $\Delta_{SC} = 6.5K$ over both ice covered water and land.

The reflectance uniformity test is applied by computing the maximum and minimum values of ASTER channel 1 or channel 3 reflectances within a 2 x 2 pixel array. Pixel array with channel 1 reflectance differences greater than 9% over land or channel 2 reflectance differences greater than 0.3% over ocean are labeled as mixed (Stowe *et al.*, 1993). The value over ocean is low because a cloud-free ocean is almost uniformly reflective, while non-uniformity is assumed to be caused by cloudiness.

Note that this test is being refined; we require that the ecosystem be the same for the pixel array. The mean and standard deviation of reflectance values for each of the 59 ecosystems will be computed for ASTER channels 1 through 3 as a function of season. It is expected that this test can be substantially improved.

The Uniform Low Stratus Test (ULST) (Stowe *et al.*, 1991) is a dynamic threshold based upon the 11 μm brightness temperature T_4. The ULST threshold Δ_{ULST} is described as:

$$\Delta_{ULST} = \exp\{A + BT_4\} - C$$

where A = -9.37528, B = 0.0341962, and C = 1.0 (oceans) and C = 3.0 (land). The constant C increases for land from the ocean value and depends on surface type. This test is applicable for the temperature range 264 K to clear-sky T_4. This test will be adapted for ASTER data; new coefficients will be determined.

A AVHRR Cirrus Test (CIRT) is applied at night over both land and ocean (Stowe *et al.*, 1991). The CIRT is defined as $(T_3-T_5)/T_5$. It also employs a channel 4-dependent threshold to implicitly account for water vapor effects. This threshold was defined by using the simulation database to plot cloud-free CIRT values against the associated channel 4 temperatures. The relatively high optical transmittance of most cirrus clouds, along with the spectrally different Planck blackbody radiance dependence on temperature, causes large (T_3-T_5) differences for cirrus clouds. The CIRT threshold is given by:

$$\text{CIRT} = -0.485 \_ 1.775 \times 10^3 T_4.$$

When $T_4 < 273$ K, this threshold is set to zero; when $T_4 > 292$ K, it is set to 0.033. This test also will be adapted for ASTER data by determining a new set of coefficients.

2.6 Texture

Texture is often interpreted in the literature as a set of statistical measures of the spatial distribution of gray levels in an image. Here it is assumed textural information is contained in the average spatial relationships that gray levels have with one another (Haralick *et al.* 1973).

The GLDV approach is based on the absolute differences between pairs of gray levels I and J found at a distance d apart at angle ϕ with a fixed direction. The difference-vector probability density function $P(m)_{d,\phi}$ is defined for m = I - J, where I and J are the corresponding gray levels, and is obtained normalizing the gray-level frequencies of occurrence by the total frequencies. From this density function, the following textural measures are computed:

mean $$\mu_{d,\phi} = \sum_m mP(m)_{d,\phi}$$

standard deviation

$$\sigma_{d,\phi} = \left[\sum_m (m - \mu_{d,\phi})^2 P(m)_{d,\phi} \right]^{1/2}$$

contrast $$\text{CON}_{d,\phi} = \sum_m m^2 P(m)_{d,\phi}$$

angular second moment

$$\text{ASM}_{d,\phi} = \sum_m \left[P(m)_{d,\phi} \right]^2$$

entropy $$\text{ENT}_{d,\phi} = -\sum_m P(m)_{d,\phi} \log P(m)_{d,\phi}$$

local homogeneity

$$HOM_{d,\phi} = \sum_{m} \frac{P(m)_{d,\phi}}{\left(1 + m^2\right)}$$

cluster shade

$$SHAD_{d,\phi} = \frac{\left[\sum_{m} \left(m - \mu_{d,\phi}\right)^3 P(m)_{d,\phi} \right]}{\sigma^3_{d,\phi}}$$

cluster prominence

$$PROM_{d,\phi} = \frac{\left[\sum_{m} \left(m - \mu_{d,\phi}\right)^4 P(m)_{d,\phi} \right]}{\sigma^4_{d,\phi}} \quad -3.$$

These textural measures are computed for LANDSAT channels 4, 5, and 6 for d=1 pixel separations and at ϕ=0° and 90°. Plots of representative cloud texture measures as a function of pixel separation distance and angle ϕ are shown in Welch *et al.* (1988) and for a variety of ice and snow backgrounds in Welch *et al.* (1990).

2.7 Artificial Intelligence Classifiers

The ASTER Polar Cloud Mask Algorithm relies upon several artificial intelligence approaches to improve classification accuracy. While these approaches require more "time" to learn, operationally they are no more cpu intensive than are traditional approaches. All of these algorithms have been written in-house; they require no commercial software.

2.7.1 Back Propagation Neural Networks

A neural network consists of objects called nodes and weighted paths connecting these nodes. Each node has an activity represented by a real number. This activity value is computed as a nonlinear-bounded monotone-increasing function of a weighted sum of the activities of other nodes that are directly connected to it.

The proposed neural network has four processing layers: 1) an input layer consisting of a node for each of the selected spectral and textural features; 2) two hidden layers, each consisting of a set of nodes; and 3) an output layer, consisting of a node for each class. The activity of node K is denoted by V_K, and a weight on a path from node L to node K is denoted by W_{kl}. Then

$$V_K = f\left(\sum_L W_{KL} V_L \right),$$

where f is a nonlinear function.

The determination of the appropriate weights W_{KL} is referred to as learning. A neural network may be viewed as a nonlinear vector-valued function: O = F(I), where O is a vector with one component for each activity of an output node, and I is a vector with one component for the activity for each input node. In the supervised learning mode, for each possible input vector I, an associated output vector O is specified. The function of the learning algorithm is to choose the value of the weights so that F(I) is a good approximation of O (Lee *et al.*, 1990). Backpropagation (Rumelhart *et al.*, 1986) refers to the process of interactively determining the weights W_{KL} that locally minimize the global error E:

$$E = \sum_{L=1}^{N} \left\| O_L - F(I_L) \right\|^2.$$

The algorithm is a special case of gradient search in which the weights are initialized as small random numbers and are repeatedly updated at the nth iteration according to the rule $\Delta W = -\eta \nabla E$, $W_{n+1} = W_n + \Delta W$, where W is a vector composed of the weights, ∇E is the gradient of the global error, and η is the learning rate. Additional details are given by Lee *et al.* (1990) and Hecht-Nielsen (1990).

2.7.2 Don't Care Neural Networks

A perceptron network consists of an input layer, an output layer and weights which define linear separating surfaces. Each pattern class C_i is separated by hyperplanes from all other surfaces. It has long been known that this network has very limited capabilities. Consider three tangent circles, each of which represents a class in 2-space. Neither traditional classifiers nor the perceptron network can find separating surfaces to correctly classify the points in the circles. However, the problem can be solved by a three layer network or by training the network to find pairwise linear separating surfaces. Training a network to produce pairwise linear separating surfaces requires that for any class C_m, the linear function corresponding to the separating hyperplane C_i/C_j will have the value 1 if m = i, a value of 0 if m = j and a "don't care" X output otherwise.

For a two-layer network, the surfaces separating the various classes are linear. Similarly, in a multi-layer network, non-linear surfaces separate the classes. Again, pairwise separating surfaces can be constructed using "don't care" outputs. In the perceptron case, the addition of "don't care" outputs broadens the repertoire of problems the network can solve. For multi-layer networks, a different benefit results. The hidden layer allows the decision surfaces to be formed into arbitrarily complex shapes. The surfaces initially are "simple", and additional training (i.e., iterations) introduces the more

complex elements into the separating surface. The network can be trained to find the simpler pairwise separator surfaces and then construct a more complicated separating surface from pieces of these simpler curves. As a result, fewer iterations are required to train the network. Our studies show that this approach can simplify the training significantly and reduce the training time by two orders of magnitude.

The steps in the algorithm can be summarized as:

Step 1: Determine the number of output nodes needed to represent the pattern classes.

Since the network will produce pairwise separating surfaces, the number of output nodes required for this technique is:

$$\binom{N}{2} = \frac{N(N-1)}{2},$$

where N is the number of classes. In contrast, traditional approaches only require N output nodes.

Step 2: Build the class representations.

Consider the desired node outputs for a class to be a bit string, where each position in the bit string serves as a discriminator between two classes. For each pair of classes, select a bit not previously chosen to be the discriminator and set that bit in one string to 0; set that same bit to 1 in the second string. After all pairs have been processed, fill the remaining positions with "don't care" symbols. This simple process can be easily automated and introduces only a small overhead penalty to the training algorithm.

For example, a 4 class problem requires 6 output nodes. Using the above algorithm, one possible assignment of output values to classes can be found in the following table.

	Bit Number					
	1	2	3	4	5	6
class 1	1	1	1	x	x	x
class 2	0	x	x	1	1	x
class 3	x	0	x	0	x	1
class 4	x	x	0	x	0	0

Possible mapping of classes to node output values

Note that bit 1 discriminates between class 1 and 2, bit 2 discriminates between class 1 and 3, and so on. The symbol "x" denotes a don't care value.

Step 3: Train the network.

During training, error is measured at the output nodes and used to adjust the network weights using back-propagation. In our experiments, the error measure:

$$network\_error = \sum_k (actual_k - desired_k)^2$$

was used. However, unlike the standard back-propagation algorithm, the above error is not calculated at the nodes which have a don't care designation. The set of weights that will be adjusted during a particular training episode is, therefore, a function of the input pattern. Note, however, that all input to hidden weights are updated.

Step 4: Classify the pattern.

To classify the pattern, simply compare the outputs to the bit strings for each class. Note that an output pattern can match at most one class since there is a discrimination bit for each pair of classes. However, it is possible that an output pattern will not match any class. As with standard back-propagation, the option exists to force a match by selecting the class to which the output pattern is in closest agreement.

2.7.3 Fuzzy Logic

Class mixtures are often classified as a single class, thereby leading to poor information extraction. This is due to uncertainty in the membership concept of the classical set theory. This representation scheme has difficulty in dealing with elements that partially belong to two or more sets. In order to improve the information representation, the concept of fuzzy set theory has been used. Fuzzy logic is concerned with formal principles of approximate reasoning; i.e., it aims at modeling imprecise modes of reasoning to make decisions in an environment of uncertainty.

The greater expressive power of fuzzy logic derives from the fact that it contains, as special cases, not only the classical two-value and multivalued logical systems but also probability theory and probabilistic logic. The main features of fuzzy logic that differentiate it from traditional logical systems are the following:

1. In two-valued logical systems, a proposition p is either true or false. In multivalued logical systems, a proposition may be true or false or have an intermediate truth value, which may be an element of finite or an imprecise characterization of a numerical truth value.

2. The predicates in two-valued logic are constrained to be crisp in infinite truth value set T. In fuzzy logic, truth values are allowed to range over the fuzzy subsets of T. Thus the fuzzy truth value may be viewed as the sense that the denotation of a predicate

must be a nonfuzzy subset of the universe. In fuzzy logic, the predicates may be either crisp (e.g., "mortal", "even") or fuzzy (e.g., "tired", "tall", "cold").

3. Two-valued as well as multivalued logics allow only two quantifiers: "all" and "some". By contrast, fuzzy logic allows, the use of fuzzy quantifiers exemplified by "most", "many", "several", and so on. Such quantifiers may be interpreted as fuzzy numbers that provide an imprecise characterization of the cardinality of one or more fuzzy or nonfuzzy sets. In this way, a fuzzy quantifier may be viewed as a second-order fuzzy predicate. On the basis of this view, fuzzy quantifiers may be used to represent the meaning of propositions containing fuzzy probabilities, and thereby make it possible to manipulate probabilities within fuzzy logic.

FUZZY MEMBERSHIP FUNCTIONS. S function or membership function: The S function is defined as follows:

$$S(x,\alpha,\beta,\gamma) \Rightarrow \begin{cases} 0 & \text{for } x \le \alpha \\ 2\left(\frac{x-\alpha}{\gamma-\alpha}\right)^2 & \text{for } \alpha \le x \le \beta \\ 1-2\left(\frac{x-\gamma}{\gamma-\alpha}\right)^2 & \text{for } \beta \le x \le \gamma \\ 1 & \text{for } x \ge \gamma \end{cases}$$

Rather than maintaining a table of data defining the membership function, the data can be easily and compactly represented by a formula. Since the polar spectral and textural features usually can be represented approximately by a Gaussian distribution, the S function is an inadequate representation of the data. A modification of this function, which is called the Π function, then is used to represent the fuzzy sets. The Π function is defined as follows:

$$\Pi(x;\beta,\gamma) \Rightarrow \begin{cases} S(x;\gamma-\beta,\gamma-\beta/2,\gamma) & x \le \gamma \\ 1-S(x;\gamma,\gamma+\beta/2,\gamma+\beta) & x \ge \gamma \end{cases}$$

In a Gaussian distribution, the spread extends to about 3σ, which contains 99% of the energy. Initially, the fuzzy sets used in our studies closely approximate the Gaussian curve. However, this can result in a number of samples being unclassified by the expert system. The spread can be gradually increased such that most fuzzy sets overlap. This improves the classification accuracy of the unclassified samples, and it does not change the accuracy of the samples previously classified.

THE FUZZY EXPERT SYSTEM (ES). A fuzzy ES includes two other elements, in addition to the components of a conventional system [Cox, 1992]: "fuzzifiers" which convert inputs into their fuzzy representations, and "defuzzifiers" which convert the output of the inference process into a single numerical value within the range of values of

the output variable. The numerical output is used to adjust the state of the system being controlled.

A fuzzy control variable may have several states, each state being represented by a membership function. Suppose we are able to classify cloud from clear land and open water by just using the albedos computed from channel one (CH1) and temperature from channel four (CH4). Figure 7 shows the different states for these two measures, with CH1 defined by the five albedo states: very low, low, medium, high and very high, CH4 defined by the three temperature states: cold, normal and warm.

The albedo measured in CH1 generally is higher for clouds than for land and water. CH4 generally is warm for land and cold for clouds. The above reasoning might lead to the following set of fuzzy rules:

Rule 1: IF CH1 is **very low** and CH4 is **normal** THEN class is **water**

Rule 2: IF CH1 is **low** and CH4 is **warm** THEN class is **land**

Rule 3: IF CH1 is **medium** and CH4 is **cold** THEN class is **cloud**

As shown is Figure 15, for a given image sample, the input value for CH1 is 0.17 and 0.4 for CH4; the fuzzifier then computes the degree of membership (DM) for one or more of these fuzzy states. In this case, the states "very low" and "low" of CH1 have a membership values of 0.5 and 0.25, respectively. The other states for CH1 are zero. Similarly, the only state of CH4 with a value different than zero is "normal", with a value of 0.60. The confidence level (CL) for each rules is computed by combining the DMs associated with each condition using the following certainty theory formula [Luger and Stubblefield, 1989]:

$$CL(C1 \text{ and } C2) = MIN(DM(C1), DM(C2))$$

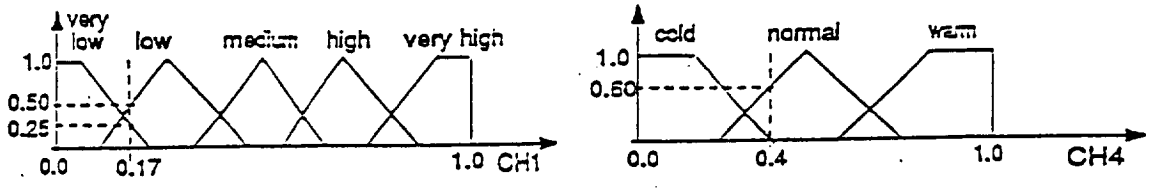


Figure 15: Fuzzy states for the Ch 1 and Ch 4 features

where C1 and C2 are the conditions of the rule. The CL for rules 1, 2 and 3 are:

rule 1: min(0.25, 0.60) = 0.25
rule 2: min(0.25, 0.0) = 0.0
rule 3: min(0.0, 0.0) = 0.0

Since rule 1 has the higher confidence level, the class selected is "water" which corresponds to the action of rule 1.

The classification process is performed with the aid of a general fuzzy expert system (GFES). GFES can handle different membership functions for describing the different states of the control variables. These functions are: triangular, trapezoidal, one-, two-, and three-dimensional normal distributions, PI function, S function, and elliptical cones. The height for all these functions is equal to 1, since any membership function can have any real value between 0 and 1. The multivariate normal distribution is a modification of the one dimensional normal distribution.

Usually, triangular, trapezoidal, PI and S functions are used by knowledge engineers for the definition of fuzzy ES's. Since our classifier uses control variables which are often assumed to belong to normal distributions, we have extended the usual set of function types to accommodate the definition of fuzzy states with one- and multi-dimensional normal distributions. Our experiments show that by increasing the number of dimensions, the classifier is able to separate better the different classes.

Three input files are required to run GFES: a control variable file, a rule file, and a facts file. The control variable file requires the following information for each control variable: the name of the variable (e.g., temperature), the type of function(e.g., 2), the number of states, the state names (e.g., hot, cold), and the values that define each state's membership function (e.g., the x1, x2, x3 values for a triangular fuzzy set).

The format for each control variable is:

```
( variable_nametype   #_of_states )
(state_name_1         value_1  value_2 ... )
      :                 :
(state_name_n         value_1  value_2 ... )
```

The rule file defines the set of rules for the expert system. Using the above set of rules, the last two rules are defined in GFES as follows:

```
( CH1      low )
( CH4      warm )
    =>
( class    land )
( CH1      medium )
( CH4      cold )
    =>
( class    cloud )
```

The first two lines of each rule represents the IF part, and each line is referred to here as a condition. The last line of a rule represents the THEN part. This line is referred to here as the action. The notation for writing a rule and facts follows the notation used by the CLIPS expert system shell developed by NASA at the Johnson Space Center [Giarratano and Riley, 1989].

The facts file consists of the values for the input control variables. These values, referred to here as the fact vector, are enclosed in parenthesis and are separated by spaces.

The data from the three input files must be read before any logical action can take place. First, the values for each input variable are converted into their fuzzy representations. That is, the degree of membership (DM) on each input variable state is computed. Next, for each rule the program computes the confidence level (CL) of the rules using the formula described above.

The CLs of two rules, R1 and R2, with the same action are combined using the following certainty theory formula:

$$CL(R1, R2) = CL(R1) + CL(R2) - CL(R1)*CL(R2)$$

The output consists in the class or classes present in the region or pixel with an associated value representing the percentage of the class within the region or pixel.

2.7.4 A Multi-Stage Neural Network Classifer

The usefulness of neural networks for classification problems is based upon a network's ability to construct arbitrarily complex decision surfaces. This frequently is accomplished by training a single network to separate all classes simultaneously. Thus, the training algorithm must find a single set of weights which accurately classifies all

samples. This is analogous to sorting a million names by moving all of them at once and hoping that the resulting ordering is closer to a sorted list. People sort large lists by first separating the elements into smaller groups, such as by the first letter in a name. The resulting sublists are then sorted. A similar approach is taken with the network structure described here. The classes are first grouped into clusters and a separate neural network, a "leaf" network, is associated with each cluster. A "switching network" is responsible for selecting the appropriate leaf network to perform the final classification task.

This approach has several advantages over the traditional monolithic training algorithm. First, the resulting network is a collection of "plug-in" components. If a more efficient switching network can be identified, that component can be removed from the tree and replaced with the new network without disrupting the operation of the remaining networks. This structure does not require homogeneous topologies or training algorithms, giving the designer flexibility to attack localized problems with appropriate solutions. Similarly, if additional data from one class becomes available, the corresponding leaf network can be removed, retrained and reinserted into the tree without affecting the remaining networks. A second advantage is that the resulting networks are smaller and thus require less time to train. In addition, since fewer separating surfaces must be identified, each network has a simpler problem to solve than a single network. This too contributes to decreased training requirements. Third, since the networks operate independently, they can be trained in parallel. A multi-processor system or a collection of workstations can be used to train the structure in approximately the amount of time required to train the largest network in the structure.

METHOD. The structure consists of four components: the switching network, the leaf networks, the clustering algorithm, and the error recovery algorithm. Note that a single decision network may not be appropriate for every problem. A hierarchy of switching networks is also implemented, and while effective, is not necessary for our problem. However, other classification tasks may require additional layers of switches.

As stated above, the leaves of the decision tree are neural networks. In these experiments all networks are trained with back propagation, but it is not necessary to do so or even to have all networks trained using the same algorithm. One of the advantages of the "plug-in" components is that any type of network can be inserted at any point in the tree.

The number of leaf networks and the distribution of classes among the leaf networks may be accomplished by a clustering algorithm. A variety of algorithms exist for this purpose. One such algorithm, found in (Duda and Hart, 1973) builds a collection of minimal spanning trees to form clusters. This suffers from a tendency to form "chains" rather than clusters and may only be appropriate as a starting point. Analysis of the data may also reveal a total lack of structure, which, while it gives no guidance on how many clusters or which classes to include in each, may indicate that the selection of these parameters is not critical to correct operation of the algorithm.

An error recovery algorithm is necessary to handle situations in which the wrong leaf network is selected by the switching network. In many situations, if a network is not able to classify a vector, it will produce small output at each of the output nodes. The current solution is to set a threshold value and if none of the outputs reaches the threshold value, the current leaf network is declared to be inappropriate and an alternate selection is made. The leaf network which receives the second largest output value from the switching network is selected and the process is repeated.

The algorithm can be described as follows:

step 1: Cluster the classes.
step 2: Train a switching network to classify a vector as a member of a given cluster.
step 3: Train each leaf network to discriminate between the classes for which it is responsible. Note that all of the leaf networks and the switching network can be trained simultaneously.
step 4: Present a testing vector to the switching network. The switching network will select a leaf network. The leaf network will classify the vector or have an insufficient response to make a classification. In that event, the switching network selects the next most likely candidate and repeats step 4.

PRELIMINARY RESULTS. This method was tested on a character recognition problem. Computer generated characters in six fonts were digitized and a feature extraction mechanism was used to create 156 vectors, each containing 14 elements (Fuji and Monita, 1971; Logar *et al.*, 1994). These were divided into groups of 78 vectors. One group contained vectors representing each character in three fonts and was used for training. The other group of 78 vectors contains the remaining three fonts and was used for testing.

The switching network used was the simplest possible: a neural network was trained using back propagation to determine which leaf network should do the classification. Four leaf networks were used, thus the "switching network" contained 14 input nodes, one 4 node hidden layer and 4 output nodes. The network was trained using back propagation and was able to identify the correct leaf network 100% of the time for the training data and 98.7% times for the testing data.

The four leaf networks were also trained using back propagation and each consisted of 14 input nodes, one 6 node hidden layer and 7 output nodes. The number of leaf networks was selected arbitrarily after clustering algorithms run on the data revealed that the vectors were approximately uniformly distributed throughout the vector space. As a consequence, no guidance was available for this problem on the number of clusters. The composition of each cluster was determined by a variation on the nearest neighbor algorithm which was modified to create balanced clusters.

Several criteria were used to compare this approach to a single back propagation network. First, consider the size of the network. A single back propagation network with 14 input nodes, 20 hidden nodes and 26 output nodes was trained. Many configurations were tried but this topology produced the most accurate predictions. For that network, 60 nodes and, including bias weights, 846 weights were required. The number of weights is particularly important since each must be updated for one iteration. Training was stopped at 700 iterations by the criteria that the change in the error over a 100 iteration period was smaller than a predefined epsilon. This resulted in a total of 592,200 weight updates. Since weight updates are the most expensive part of the algorithm, this is a good measure of relative speed. In contrast the decision tree neural network with its five networks contained 130 nodes and 636 weights. However, since each leaf network is assigned a simpler task, fewer training iterations were required. The average number of iterations for all five networks was 400, resulting in 254,400 weight updates. In this case, the decision tree neural network required approximately 25% less storage for weights and reduced the number of weight updates by approximately 42%. Timings were also conducted using a 486-based PC. The single network required 5.39 seconds/iteration to train, or a total of 3773 seconds. Two timings must be considered for the decision tree network. First, code was written to train the networks simultaneously on a single processor machine. The total time was 402.8 seconds. However, one of the advantages of this architecture is that all five networks can be trained in parallel. Thus, a five processor machine, or five processes running on five dedicated workstations, can produce the weights in approximately 168.4 seconds, or the maximum of the five independent training times. Thus, training times were reduced by 89% for the single processor implementation. Note that the number of weight updates is reduced by 42% while training time is reduced by 89%. This difference can be attributed to the fact that the amount of time required for a weight update is dependent upon the size of the network.

Classification accuracy was also measured. The single network had an accuracy of 100% on the training vectors and 82.7% on the testing vectors. The multi-stage network also classified 100% of the training vectors and 87.2% of the training vectors correctly. Thus, performance was not affected and the time and space required to achieve this performance were significantly reduced.

FUTURE WORK. The experiments reported here are limited and much more work must be done to fully develop these ideas. Each of the four components must be more deeply studied. The switching network has been designed to allow the incorporation of "hints", information external to the vector to be classified. At present, the hints are given a numeric value and simply incorporated into the switching node output. The mechanism has been developed, but not yet tested, to allow the incorporation of a fuzzy logic and rule based system to provide the hints to the switching network. The rule based system may produce its candidate for which leaf network should do the classification. That information, in conjunction with the output from the switching network, can be used in a fuzzy logic algorithm to determine which leaf network to use.

One obvious addition to the leaf networks is the incorporation of the "don't care" training algorithm (Logar *et al.*, 1994; Watters, *et al.*, 1993). In a don't care network, the separating surfaces are combinations of surfaces which separate pairs of classes rather than the traditional approach of separating one class from all others. The algorithm is described above and has proven very effective in a variety of applications. As with the structure described above, don't care networks build complicated separating surfaces from simple components, each of which is easier and quicker to identify than the single separating surface.

Only the simplest clustering techniques have been considered here. Additional work must be done to determine a suitable algorithm or to identify criteria for selecting from a collection of clustering algorithms. One approach that has yet to be explored is the use of a genetic algorithm for forming clusters. Such an algorithm would start with random clusters and use genetic operators to "breed" better solutions, that is, better clusters.

Finally, a better error correction algorithm is required. The method described above is not effective if the switching network makes a mistake, sends a vector to the wrong leaf network and the leaf network claims to strongly recognize it as a member of an incorrect class. However, the effectiveness of the existing mechanism may be improved by either an alternate clustering algorithm or by the incorporation of the rule-based hints and fuzzy logic switching system.

3.0 ALGORITHM DESCRIPTION

In this section some required preprocessing steps are described as well as the algorithm itself. Two types of preprocessing are required. First, preprocessing of each scene is required to provide a one to one spatial relationship among the bands in the 3 ASTER sensors (i.e., VNIR, SWIR, and TIR). In addition, some normalization is performed. Second, prior to making the algorithm operational (or during subsequent evaluation periods), various functions, weights, parameters need to be defined based on training samples, empirical evidence, and a priori knowledge. Next, an overview of the algorithm is provided, followed by a more detailed description of the algorithm classification strategy.

The major algorithmic steps include (Fig. 16):

(1) Preprocessing
(2) Table Lookup Classification
(3) Fuzzy Expert Classification
(4) Spatial Context Tests
(5) Quality Assurance Tests

Algorithm Flow Chart

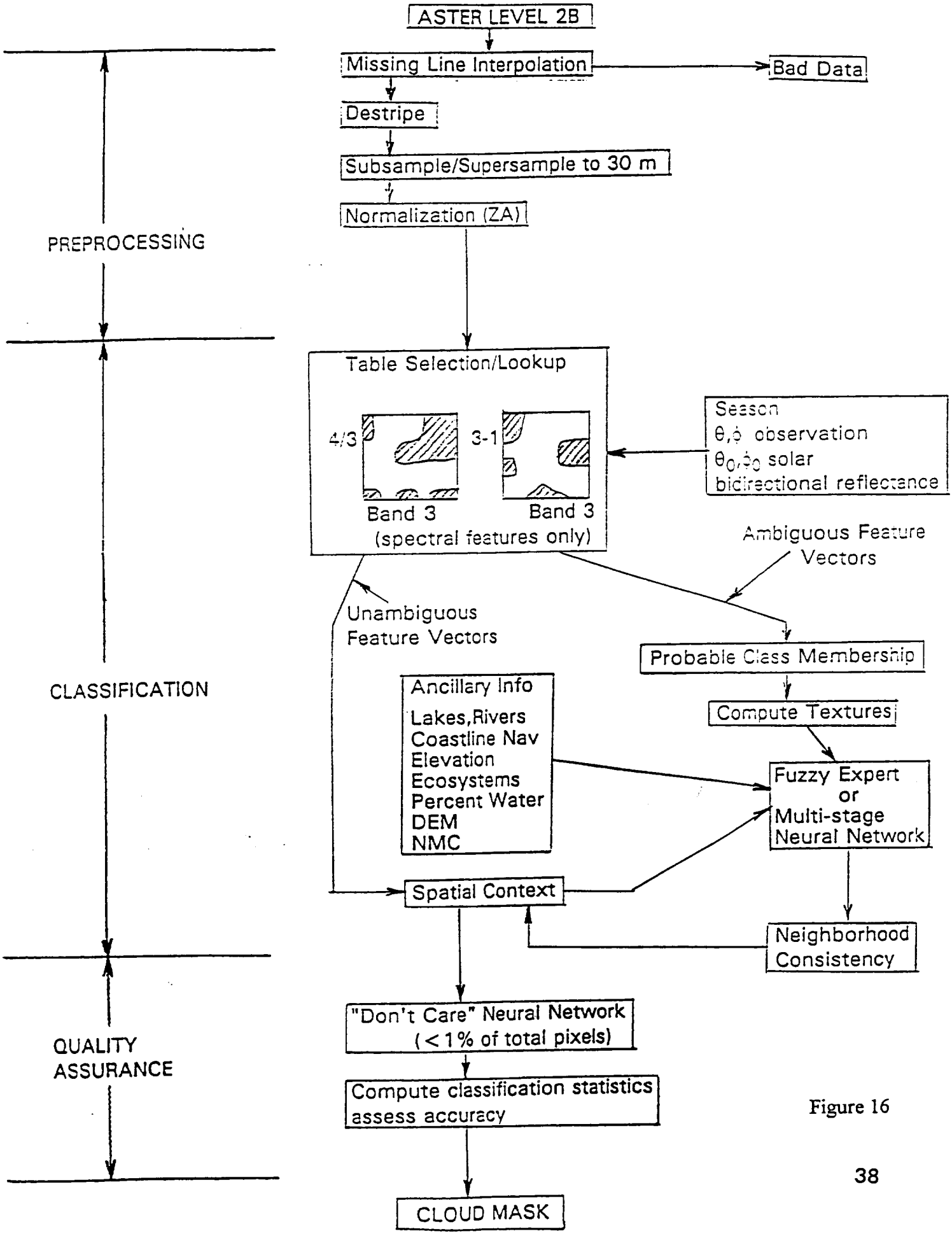


Figure 16

3.1 Preprocessing

A number of preprocessing steps are made to the ASTER data before the cloud masking algorithm is applied. These preprocessing steps are described below:

(1) The AVHRR data will be navigated using the World Data Bank II coastline system at 500 m resolution. The world is divided into five regions, each of which contains the following: coastlines, islands, lakes, reefs, salt pans, ice shelves, glaciers, rivers, railroads, international boundaries, and internal political boundaries. An example of navigation accuracy is shown in Fig. 17 for AVHRR data.

(2) The NAVY 10 minute database is a 1080 x 2160 byte array covering 180 degrees in latitude from north to south pole and 360 degrees in longitude. The global elevation map is shown in Fig. 18. The surface elevation characteristics are:

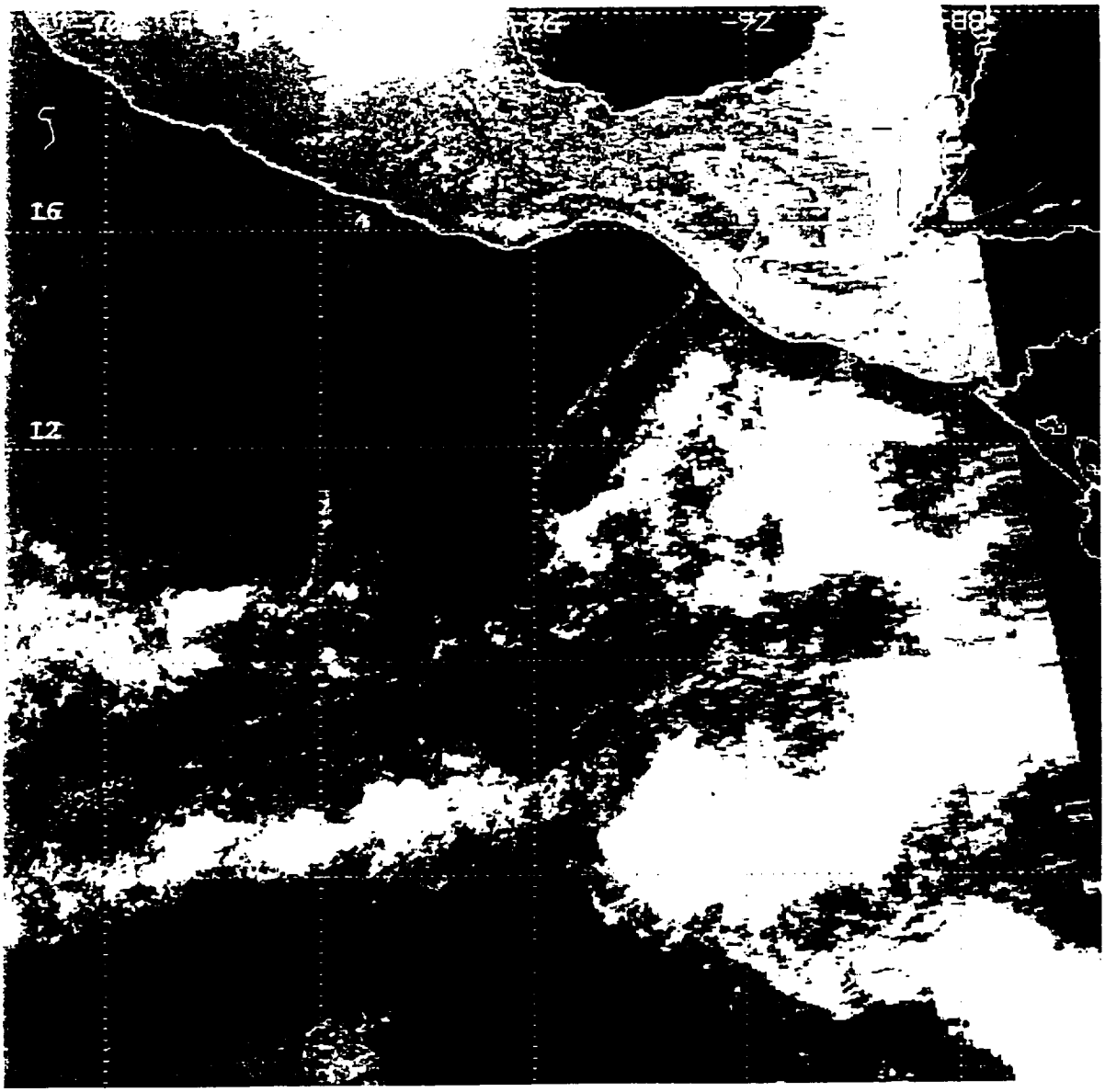
Table 4

Codes	Feature
0	salt or lake bed
1	flat or relatively flat
2	desert (or for high lat, glaciers or permanent ice)
3	marsh
4	lake country or atoll
5	major valleys or river beds
6	isolated mountains, ridge or peak
7	low mountains
8	average mountains
9	extremely rugged mountains
62	ocean

Note that multiple characteristics are defined in this system; an example is code 14 = flat lake country or atoll. In addition, this code contains the percentage (an integer between 0 and 100) of water in the 10 minute box. The global elevation map is shown in Figure 19.

(3) The EPA Global Ecosystems (WE1.4D) Database also is a 1080 x 2160 byte array which contains 59 different ecosystems classes (Fig. 20).

(4) The U.S. NAVY/NOAA Sea Ice Product provides weekly reports of fractional ice coverage at spatial resolution of about 18 km.



Navigated AVHRR GAC image

(88 to 104 W longitude and 2 to 20 N latitude)

Figure 17

Figure 18

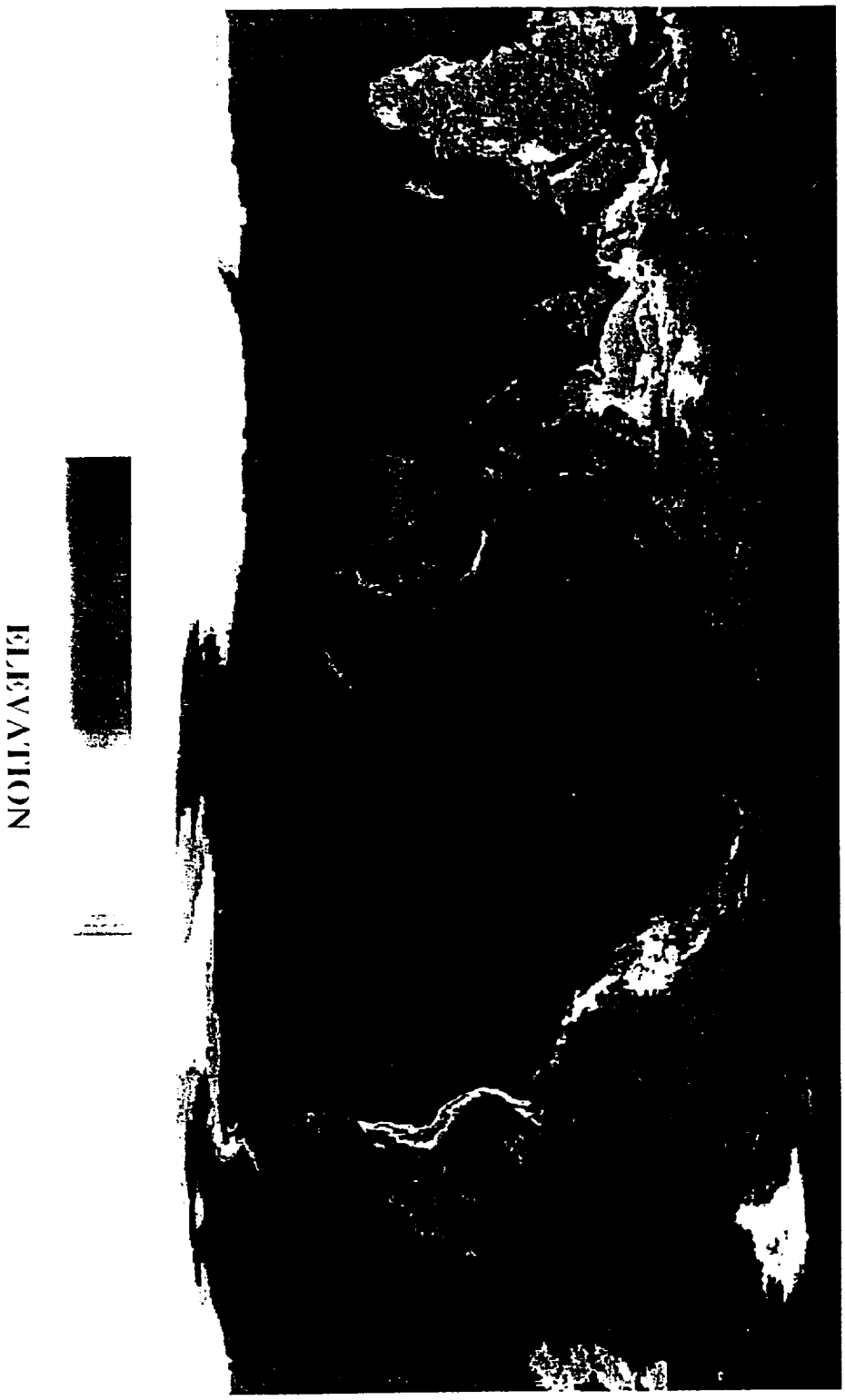


Figure 19

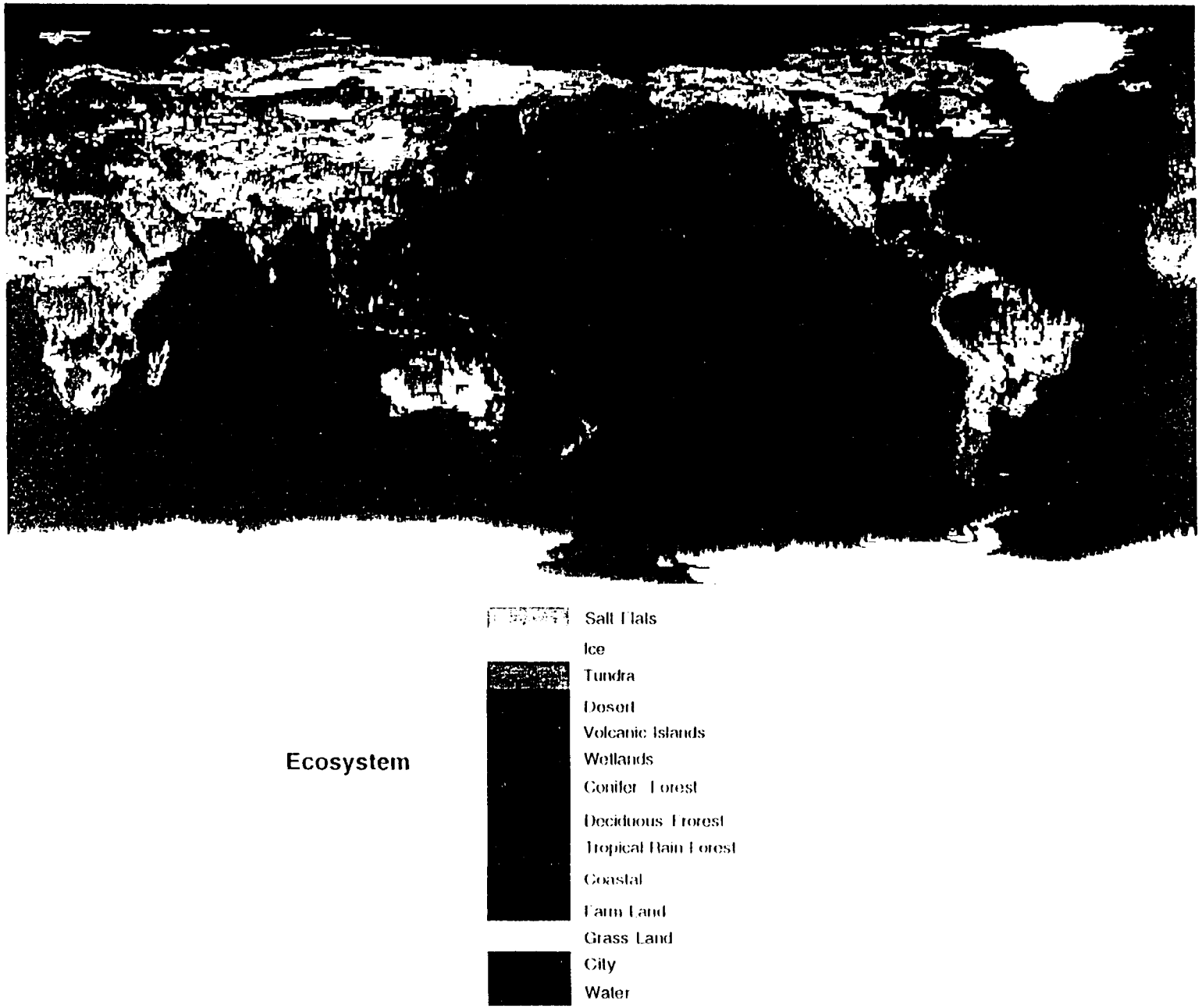

Ecosystem
Salt Flats
Ice
Tundra
Desert
Volcanic Islands
Wetlands
Conifer Forest
Deciduous Forest
Tropical Rain Forest
Coastal
Farm Land
Grass Land
City
Water

Figure 20

(5) The NOAA Snow Data Product provides weekly report of snow cover at a spatial resolution of 150-200 km; snow is reported if the grid cell is more than 50% covered.

(6) The NMC 3-hour surface analyses of temperature and wind speed.

(7) The MODIS daily snow/ice output (available after launch).

First the scene will be navigated, with coastlines, oceans, lakes, rivers, marshes, reefs, permanent ice regions, deserts and salt beds noted. Second, each land pixel will be designated as relatively flat, valley, isolated mountainous regions, low mountains or hills, average mountains, or high mountains. From the NOAA Snow Data Product each land pixel will be designated as probably/probably not snow covered. Each land pixel also will be classified as to its ecosystem, along with a more general ecosystem classification of urban, forest, woodland, grassland, shrubland, tundra, arid vegetation and highland vegetation. Ocean regions will be classified as water, coastline (including islands), possibility of isolated icebergs, marginal ice zone, and nearly solid ice (leads may be present).

3.1.1 Image Navigation

The navigation of satellite imagery is an important preprocessing procedure in digital image analysis. The latitude and longitude of each pixel must be known to the desired accuracy if data from two sensors are to be matched. The distortions due to earth shape, earth rotation, variations in satellite orbit and satellite altitude must be accounted for in any navigation procedure. There are two types of image navigation. The first is called direct image referencing where the geographic grid is distorted to match the satellite image projection and the second is called inverse image referencing where the image is corrected and resampled to fit a desired geographic map projection. The second method is widely used in the remote sensing community, and several navigation packages are available (e.g. Emery *et al.*, 1989). The inverse image referencing can be further divided into two methods. In the first method, known ground control points (GCPS) can be used to correct for errors in earth shape, scan geometry, satellite orbit and satellite altitude. This method relies on known geographical features in visible imagery and is often time consuming because the user must interact with the image and locate the GCP's. Also, this method cannot work over the open ocean because there are very few GCP's. The second method is to utilize the satellite ephemeris data (orbital parameters) and locate the satellite as a function of time. This method is less time consuming and no user intervention is necessary. One common source of error in this method is the inaccurate timing at the ground station where the data are collected. The ephemeris satellite navigation techniques can be used to produce navigated images with an accuracy of a single pixel (Ho and Asem, 1986).

Over land, it is often possible to locate geographical features in visible imagery and utilize a map overlay in order to test the accuracy of the navigation procedure. The

corrections for any shifts in the image are termed "nudges" by users. In this procedure, the whole image is moved up or down until the geographic features in the image line up with the map overlay. This offset can then be applied to the latitude and longitude of each pixel. Over the open ocean, it is often difficult to detect geographical features. However, high spatial resolution cloud features can be used (Feind and Welch, 1994).

We have navigated several AVHRR LAC and GAC images over land and open ocean. An example is shown in Fig. 17. This same procedure will be used on the LANDSAT and ASTER data. Errors in navigation of about 500 m are expected.

3.1.2 Preprocessing of Image Data.

The ASTER data are obtained at 3 different spatial resolution (i.e., VNIR-15 m, SWIR-30 m, TIR-90 m). The classification is derived at 30 m spatial resolution. Therefore, after the data are tested/corrected for missing lines/columns and striping, the VNIR data is subsampled, by half, to 30 m pixel spacing. The TIR data is supersampled, by 3, also to 30 m pixel spacing. The SWIR data remains unaltered. The VNIR and SWIR band DNs are then normalized for solar irradiance, solar zenith angle, observation angle, and calibration coefficients. The TIR band DNs are converted to temperature.

3.1.3 Preprocessing for the 'Don't Care' Neural Network.

A 'don't care' neural network is used in the final stage of the algorithm as a quality assurance tester. The structure of the network is as described below. Before launch or before the algorithm becomes operational, feature vectors comprised of the spectral and textural features (Welch *et al.*, 1990) for each training sample are constructed. The neural network is then trained and tested on these samples. The weights from the training are then used in the operational neural network. We expect that multiple sets of weights will be required, based on time of day, solar azimuth, bidirectional reflectance, etc.

3.1.4 Preprocessing for the Fuzzy Expert.

1. "Filtering" - The feature vectors of the selected samples from the images pass through a "filter" process which removes samples with feature values that exceed four standard deviations from the mean. Four standard deviations is used to include approximately 99.74% of the "good" samples with feature values within four standard deviations of the mean.

2. "Normalization" - The filtered vectors are normalized using the sigmoid normalization process which maps each value from the feature vector into the interval (0,1) {Logar, *et al.*, 1992). This normalization technique transforms the mean of each feature 0.5. That is, the feature vector elements are clustered around the value 0.5.

3. "Training and testing sets" - Two-thirds of the data set from the spectral and textural database, which replacement, are used as the training data for the classification

process. The other one third is used as the test data. The term "replacement" means that each sample selected as training sample may be selected more than once. This insured an unbiased estimate of classification accuracy. The training sample is selected randomly. Training and test samples are separated for each class. The accuracy of the classifier is determined by computing the percentage of correct identification of the test samples for each class.

4. "Statistical parameters" - The mean, standard deviation and covariance matrix are computed for all classes and for all features, using the samples in the training set.

5. "ES rules and control variable files" - The control variables and fuzzy rules files are generated using the statistic parameters computed in the previous step. Our fuzzy ES has a rule for each class. Each rule has the same number of conditions and only one action.

6. "Tuning" - Conditions that do not contribute to the accuracy of the classification are removed from the fuzzy rules. This process is accomplished by first generating sensitivity reports and then selecting features that better discriminate the classes. In the next section we discuss several experiments perform in order to define the appropriate set of rules. We expect to automate this step in the near future.

3.2 Algorithm Overview

The methodology implemented in this algorithm can be characterized as hierarchical or multi-stage, as opposed to flat or single layer. The intent is to link a multiplicity of techniques in such a way that efficiency and speed are optimized while not compromising classification accuracy. Some class members are classified at a high level of confidence using a small set of spectral features using simple decision surfaces while others require larger feature sets (comprised of both spectral and textural measures) using more complex classification strategies such as fuzzy logic and neural networks. To the maximum extent possible the classification strategy is derived in knowledge of physical phenomenology. Parameterization is used only as necessary.

The algorithm has five stages or levels (see Figure 1). The first stage is where feature vectors close to class cluster centers are conservatively classified, with small computational expense, through the use of table lookups. The samples not classified by the first stage then are passed to the second stage. In the second stage additional features are computed and a fuzzy expert is used to classify ambiguous feature vectors. In the third stage a consistency test between the first and second stages is performed and reclassified pixels are as unknown if they are inconsistent. In the fourth stage, a spatial context test is performed on the samples reclassified as unknown from the third stage. And finally, in the fifth stage, a quality assurance test is performed, in which a neural network is applied to a random sampling of classified feature vectors. The scene is flagged for human expert evaluation if a simple statistical threshold is not achieved.

3.3 Algorithm Description

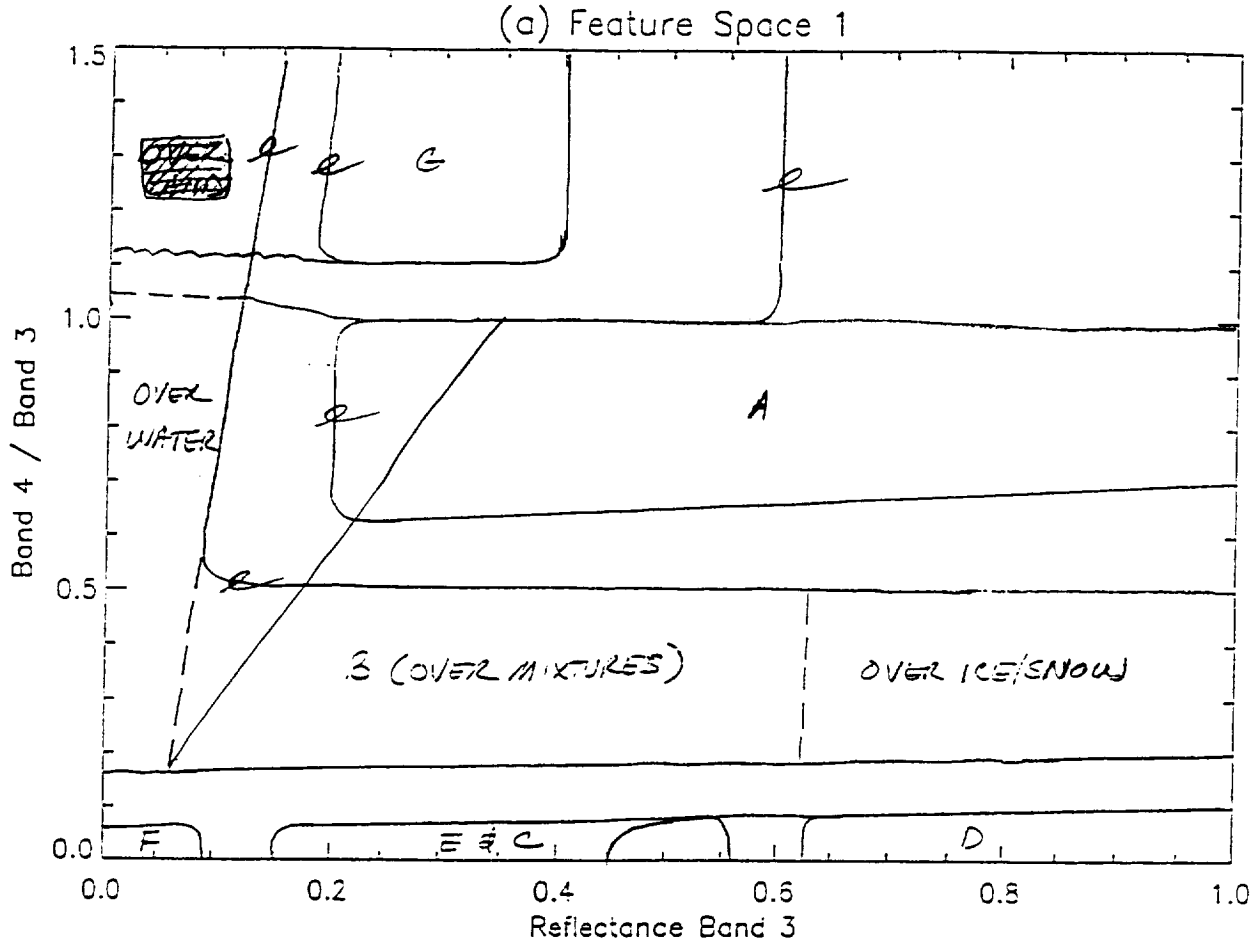
3.3.1 First Stage-Classification of Regions with Feature Vectors near Class Cluster Centers.

The first level or stage of processing is based on a set of lookup tables, each constructed in 2-dimensional feature space. Table lookups allow us to classify large sets of pixels efficiently. For example, computationally expensive techniques such as minimum distance classification (which requires Euclidean distances), or Maximum Likelihood classification (which requires covariance matrices and inverses) are not required. In addition, conditional thresholds (conditional on other features) are possible and amorphous (non-circular, non-Gaussian) decision regions can be constructed. Currently, the size of the tables corresponds to the radiometric resolution of the LANDSAT sensors (i.e., 0-255). The size of each table is 256^2 or 64 K. (Note: Tables in which the ASTER TIR channels are used as features might require a larger range of indices to retain the 12-bit dynamic range of the data.) The size of the tables potentially can be reduced pending further testing and evaluation. Each table value has a one-to-one correspondence with a specific range of values for a pair of features. Each table location is coded with an unambiguous or ambiguous classification.

The algorithm currently uses 2 tables for a particular scene. Additional tables will be constructed as we learn more about the causes of class cluster displacement for a given class from scene to scene. Such behavior is related to one or a combination of the following: seasons, time of day, solar azimuth, bidirectional reflectance, synoptic conditions. The boundaries between classes in the 2D feature spaces have width. The width is determined empirically such that only feature vectors near class cluster centers are unambiguously classified. The lookup table boundary regions contain codes indicating the probable classes or mixtures of classes for those pixels. These boundary pixels are deferred to the next level of processing. An illustration of this first level of processing is depicted in Fig. 21a,b. Here we show Band 4/Band 3 versus Band 3 and (Band 3-Band 1) versus Band 3 of ASTER. These two 2D feature spaces have been partitioned into unambiguous classes and ambiguous boundaries between classes.

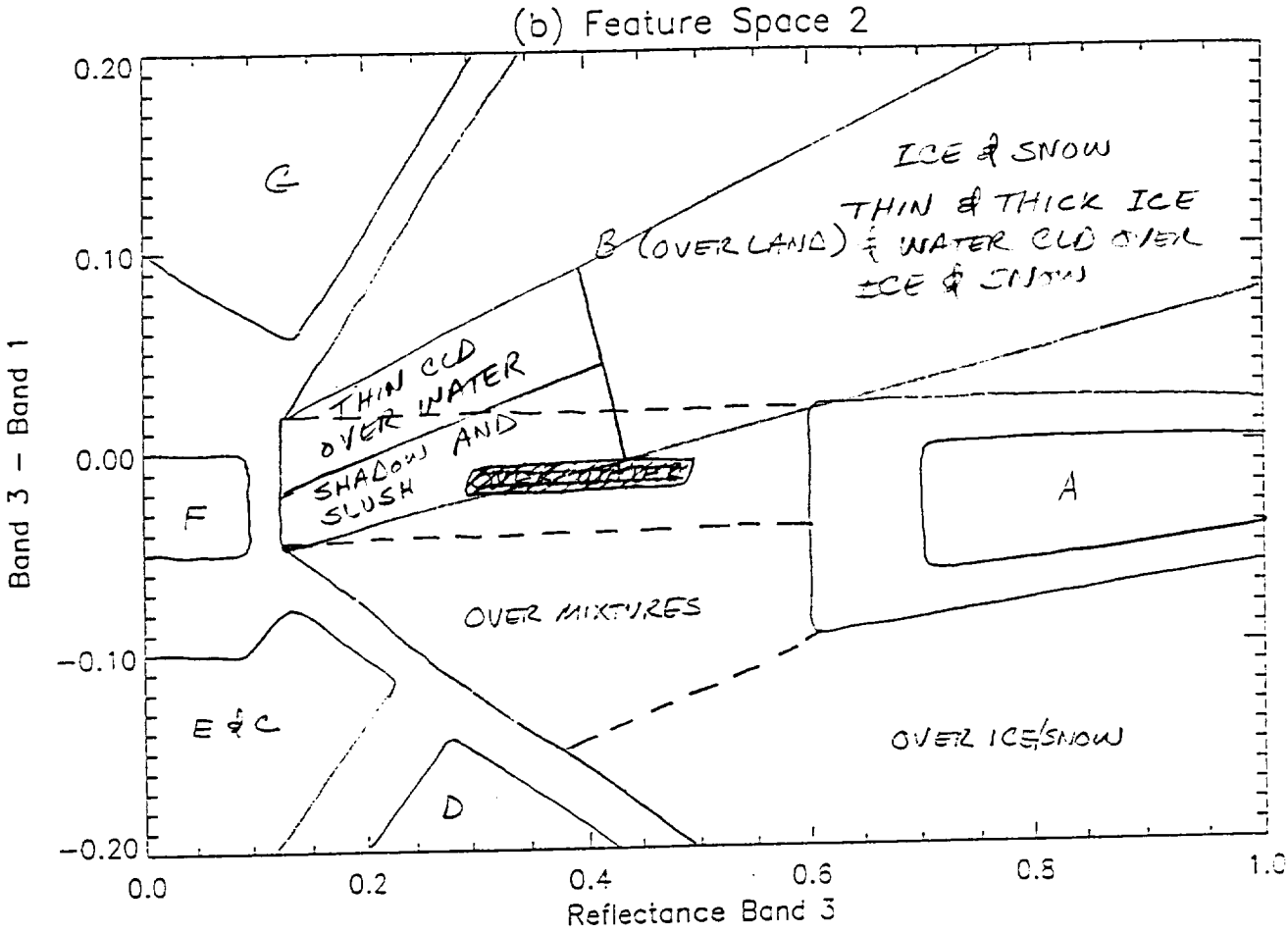
The spectral region of ASTER Band 4 is 1.6 - 1.7 μm. Solar radiation in this region is strongly absorbed by snow and ice, relative to clouds, due to their larger particle sizes and smaller single scattering albedo (e.g. see Fig. 22). Ice cloud absorbs more strongly than does water cloud but less than snow/ice. Therefore, Band 4 is very useful for discriminating any kind of cloud (thin, thick, water cloud, ice cloud) from water, snow, or ice backgrounds. The reflectance of polar land surfaces can be brighter than ice in Band 4, but another feature (Band 3-Band 1) is used to distinguish land surfaces from snow and ice surfaces.

We use the ratio of Band 4 to Band 3 because it is more strictly a function of particle size than is Band 4 alone. Band 3 is an ASTER VNIR channel at 0.76 - 0.86 μm. Water has low reflectance (<0.1) in this spectral range. Shadows on ice/snow surfaces are



A - Thick Cloud
B - Thin Cloud
C - Shadow
D - Ice and Snow
E - Wet Ice
F - Water
G - Land
Unlabeled regions ambiguous

Figure 21a: Partitions for table lookup classification in Band 4/Band 3 versus Band 3 future space.



A - Thick Cloud
B - Thin Cloud
C - Shadow
D - Ice and Snow
E - Wet Ice
F - Water
G - Land
Unlabeled regions ambiguous

Figure 21b: Same as Figure 21a except for Band 3-Band 1 versus Band 3 feature space.

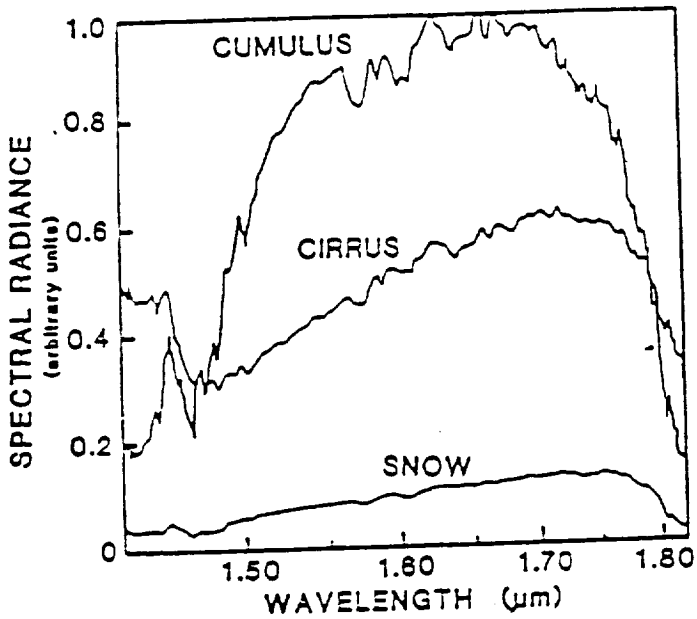


Figure 22: Relative spectra of water cloud, ice cloud, and snow in 1.6 μm spectral region.

somewhat brighter and unshadowed ice/snow are very bright. In Fig. 4 the distribution of water/shadows/ice/snow can be seen in the Band 4/Band 3 versus Band 3 feature space. They all very nearly lie on line from water to highly reflecting snow. In the plot, the pixels to the right of water and left of snow/ice can be shadowed snow/ice or mixtures of any combination of water, shadowed snow/ice, and snow/ice.

We derive the utility of ASTER Band 3-Band 1 from Li and Leighton, 1991. (The spectral range of Band 1 is 0.52-0.60 μm.) Land surfaces (including both soil and vegetation) are brighter in Band 3 than Band 1, resulting in a positive difference for land surfaces. In addition, clouds manifest very small differences in brightness as their radiative properties are very similar in these 2 spectral regions. Ice/snow surfaces are generally brighter in Band 1 than in Band 3 due to Rayleigh scattering; therefore, ice/snow surfaces result in negative values for Band 3-Band 1. In this feature space we can see that clear pixels over land lie on a straight line between water and the brightest land feature. The pixels just to the right and above water are shadowed land and/or mixtures of any combination of water, land, and shadowed land. Similarly, the same linear relationship between water and bright snow can be seen (and as also seen in the Band 4/Band 3 versus Band 3 feature space). Classifying pixels in these 2 feature spaces will produce redundant results. Currently, we use the Band 3-Band 1 feature to classify land and the Band 4/Band 3 feature to classify snow/ice. However, as we continue to test the algorithm, we will compare classifications and determine the accuracy of each. Eventually, they will probably be used in an either-or scheme. The cloud partitions in the two feature spaces are also redundant. Presently, we don't know which one provides the most accurate classification. The feature vectors located in unambiguous feature space are assigned high confidence values for classification, and their corresponding bit maps are constructed. An additional threshold test for water is applied to the ASTER TIR Band 13 and ASTER VNIR Band 4. Any regions greater than the Band 13 threshold (271°K) and less than the Band 4 threshold (0.1) are classified as water. The remaining pixels whose feature vectors lie in boundary regions will then be passed on to the next processing stage.

3.3.2 Second Stage - Fuzzy Expert Classification.

The textures described in Welch *et al.*(1990) are computed for each of the pixels with ambiguous classification from stage 1. Pixel arrays are used for the neighborhoods; the neighborhood is centered over the pixel to be classified. Then the spectral and textural features are input to a fuzzy expert system. The decision rules for each of the classes and the set of IF-THEN tests for each rule are determined during processing. (Note: The current structure of the fuzzy expert requires one IF-THEN test for each feature in each class. Some preliminary results indicate that higher classification accuracies can be achieved by reducing the number of tests in some rules. Additional study will determine whether any feature reduction can be implemented in the cloud mask algorithm fuzzy expert.) The fuzzy expert then computes a certainty value for each class. If a binary classification scheme is used (i.e., 1 bit/class) then the bit for any class with a certainty value greater than 0.35 is set to 1. If a 2 bit (or 4 level) classification scheme is used, then the 2 bits are set as follows:

Certainty Range	Classification Bits
<0.09	00
0.10 - 0.34	01
0.35 - 0.89	10
0.90 - 1.0	11

The fuzzy expert also tests the consistency of the classification with ancillary information such as coastline navigation, ecosystem maps, and surface temperatures.

3.3.3 Third Stage-Classification Consistency Test.

In the next stage, a stage 1 classification consistency test is performed. Each ambiguous pixel from the first stage is coded with a value indicating the possible classes for the pixel. If the fuzzy expert selects one of the possible classes then the classification process is terminated for that pixel. If not, the pixel classification is labeled as unknown.

3.3.4 Fourth Stage-Spatial Context Test.

For any pixel with an unknown classification, one final test is performed. The spatial context is tested by examining the 8 nearest neighbors. If 6 or more of the nearest neighbors are the same class then the pixel is assigned to that class by setting the 1 bit code to 1 or the 2 bit code to 11. If 6 or more pixels are from 2 of the same ambiguous classes assigned in stage 1, then the two 1 bit codes corresponding to the 2 classes are set to 1 or the 2 bit codes are set to 01 and 01.

3.3.5 Fifth Stage-Quality Assurance Test.

Finally, as a quality assurance test of the entire scene classification, approximately 1000 pixels are selected at random in which the spectral and textural features are computed and supplied to a trained don't care neural network. A statistic is computed for the fraction of classification agreements. If the statistic is less than 0.85 the scene flagged as suspect and deferred to a human expert for evaluation.

An alternative to the second level of processing (i.e., the fuzzy expert) is based on the preliminary study described previously. The multi-stage neural network classifier has yet to be tested on polar feature vectors. If this technique is found to be at least as accurate as the fuzzy expert and faster, it will supplant the fuzzy expert as the second level classifier. The fuzzy expert would then be substituted in the fifth stage as the quality assumer.

3.4 Variance and Uncertainty Estimates

This section is based upon our previous experience of classifying polar regions using AVHRR LAC data.

Using a combination of eight spectral and textural features derived from AVHRR polar scenes, six artificial intelligence classifiers were used to generate results for 10 classes: water, solid sea ice, broken sea ice, snow-covered mountains, stratus over ice, cirrus over ice, stratus over water, cumulus over water, multilayer cloudiness, and land. While there were other distinguishable classes present in the scenes, there were insufficient samples of these classes to present to the classifiers. Results are shown in Table 2.

Two thirds of the data set from the spectral and textural data base, with replacement, were used to generate the training data for each classifier. The remainder was used as test data to determine classifier accuracy. The term "replacement" means that each sample may be selected more than once. This insures an unbiased estimate of classification accuracy.

To generate the results, ten realizations were computed. That is, the training data were randomly selected on the basis of a random-number seed. A different random-number seed leads to a different selection of training data and somewhat different results. To compute the theoretical accuracy of the results, it is necessary to average over multiple realizations. Computations made for ten to a hundred different realizations have been made. It was found that the theoretical accuracy of the classifier can be measured sufficiently well with only ten realizations.

This theoretically sound bootstrap approach treats the simulation on a nonparametric basis. Estimates of both the correct classification probability μ and its standard deviation σ are shown in Table 2 for each of the six classifiers. Results are shown for each of the 10 classes and for the total. A total of 10 bootstrap sample sets were used in each case.

A similar approach is being used with the LANDSAT TM data. Two thirds of the data will be selected as training data, with replacement. The remainder will be used as test data. Ten realizations will be made in order to calculate the theoretical accuracy and variance of the classification results. We estimate that accuracies of 95% to 98% can be achieved for the seven classes using the ASTER polar cloud masking algorithm. This is in large part due to the high spatial resolution available which allows for positive identification of features using IVICS.

3.5 Practical Considerations

3.5.1 Numerical Computation Considerations

The majority of this algorithm development work is being done in Version 1 using IDL. Due to its ease of use and its display capabilities, IDL is the language of choice for most of our staff. However, we are making every effort to streamline and optimize the code. At present, the Version 1 algorithm to be delivered to the DAAC contains elements of IDL, IMSL, and C++. A C++ Class Library is being constructed to simulate several of

TABLE 2. Overall Classifications Accuracies for the Six Classification Methods (from Tovinkere *et al.*, 1993)

Class	FFBP		PNN		HYBRID		"Don't Care" Perceptron		"Don't Care" Back Propagation		Fuzzy Logic	
	μ	s	μ	s	μ	s	μ	s	μ	s	μ	s
1. Water	100.0	0.0	100.0	0.0	99.5	1.0	100.0	0.0	100.0	0.0	99.2	1.1
2. Solid sea ice	97.9	1.5	98.2	1.2	94.1	5.2	97.3	1.5	98.8	1.4	93.1	3.7
3. Broken sea ice	98.6	1.1	99.6	0.5	90.4	6.7	97.9	1.4	99.7	0.7	99.4	0.7
4. Snow-covered mountains	99.6	0.8	99.7	0.5	79.2	13.1	100.0	0.0	99.8	0.5	97.5	2.4
5. Stratus over ice	89.8	3.2	81.7	2.7	80.4	6.3	87.6	4.6	93.5	3.5	91.5	2.9
6. Cirrus over ice	99.2	0.9	99.4	0.9	88.1	4.3	98.0	1.2	99.6	0.4	95.1	1.6
7. Stratus over water	97.4	1.6	97.7	1.7	78.5	9.6	94.6	2.7	99.2	0.6	87.9	3.6
8. Cumulus over water	99.1	0.6	99.8	0.4	94.3	4.5	99.9	0.3	99.4	0.8	99.5	0.5
9. Multilayer cloudiness	95.8	1.5	93.9	2.4	70.4	7.0	92.3	1.8	96.7	1.7	93.2	2.1
10. Land	100.0	0.0	100.0	0.0	99.5	0.8	100.0	0.0	100.0	0.0	100.0	0.0
Overall	97.8	0.4	97.2	0.4	86.3	1.6	96.6	0.6	98.7	0.3	95.1	0.7

the IDL and IMSL subroutines. Timing tests between IDL, IMSL, and C++ have revealed that many of the IDL and IMSL calls are unnecessarily CPU intensive. We will replace inefficient IDL and IMSL function calls with the C++ Class Library functions. This approach allows the research staff to concentrate their efforts on algorithm development using the most convenient IDL and IMSL function calls without having to worry about the CPU expense.

With regard to the AI classification algorithms, it is uncertain at this time which offers the best combination of run time and accuracy. A thorough intercomparison of the AI classifiers and traditional methods, such as parallelepiped, minimum distance, Mahalanobis, and maximum likelihood, is planned in FY94-95. The purpose of this exercise is to document the most accurate and computationally efficient algorithms. Other techniques also are being investigated. In one case, easily identified pixels are being classified using both thresholding and the parallelepiped method; then, for those pixels unclassified in the first step, a succession of ever more CPU intensive classifiers will be applied. The goal is to determine if a classification hierarchy produces faster and more reliable results. To date, the AI approaches produce much higher accuracies with run times which are comparable or even faster than the traditional methods.

The speed of the algorithm is scene dependent. Scenes completely obscured by cloud, for example, and that do not contain missing lines/columns or have striping, will require no more than a minute on machine running at 6-8 FLOPS. Compiled scenes that contain significant areal fractions of thin cloud could require several stages of processing. Without further experience it is difficult to estimate a worst case or average time to process one scene. It is important to remember that most polar scene classification studies to date have been conducted on low resolution imagery using only neighborhoods of pixels, not every pixel in a scene.

3.5.2 Programming/Procedural Considerations

Much of this code is being developed using C++. However, C++ has no ANSI standards at this time. Therefore, a very conservative application of C++ functions is planned, to assure portability to any platform. We will test the code on Silicon Graphics, SUN, and IBM RISC machines. To meet the PGS Toolkit specifications for Version 2, conversion to C will be made.

Note that all of the AI classifiers have been developed in-house. Therefore, there is no reliance upon other commercial software packages. We will use the PGS Toolkit once it is made available. However, at present, all code is being developed by the staff.

3.5.3 Calibration and Validation

Two approaches will be used to validate the output of this algorithm (ie., the classification of the ASTER imagery at the spatial resolution of the SWIR sensor - approximately 30 m). The first will occur before launch and will take advantage of data

collected during the polar FIRE III IFO program which may be conducted in 1995-1996. A complete complement of data obtained from satellite sensors, airborne sensors (in situ, imaging spectrometers, profilers, lidar, etc.) and surface observations (including radar, balloons, human observations, etc.) will provide for unambiguous classification of the region under study. The algorithm will then be exercised on imagery obtained from the MAS and LANDSAT TM (and potentially AVIRIS and TIMS). The results will then be compared to determine classification accuracy.

Another approach will take place after launch. Periodically, ASTER scenes will be selected at random and a human polar scene classification expert will conduct a manual classification. We expect initially that several scenes per month will be validated and as the algorithm is adjusted and as greater confidence or adequate classification accuracies are achieved (>90%), the frequency of manual validation will be reduced to a few per month.

Between now and the time of launch, the algorithm will be tested on any high spatial resolution imagery that becomes available (e.g., from MAS, AVIRIS, TIMS, and LANDSAT TM). A human expert will be evaluating and validating the algorithm's classification results.

3.5.4 Quality Control and Diagnostics

Our current plan for assessing the quality of the output is to randomly select a predefined number of pixels from each scene (eg., 1000 from a 2000 x 2000 pixel scene) and process them with a 'don't care' neural network or fuzzy expert using all spectral and textural features. If the classification agrees, the quality of the classification will be deemed to be good. If they disagree, the quality will be flagged as suspect. Statistics for this comparison will be accumulated and some percentage classification agreement threshold will be set (eg., <85%) which will then flag the classification of the entire scene as suspect and requiring human intervention or evaluation.

3.5.5 Exception Handling

3.5.5.1 Missing Scan Lines/Columns.

The module for handling missing scan lines/columns is performed as a preprocessing task prior to generating the cloud mask. We expect that the VNIR and SWIR will potentially manifest missing columns since they are pushbroom sensors and the TIR will manifest missing rows since it is a scanning sensor. The module implements 2 functions which are detection and reconstruction/rejection. Of course, if no missing scan lines/columns are detected than no correction is applied. If the scan line/column criteria is not met (see the next section), the data is rejected for further processing; the cloud mask bit map will indicate that the data is of inadequate quality for classification.

3.5.5.2 Rejection Criteria for Missing Scan Lines/Columns

If 3 or more consecutive scan lines/columns are detected than those 3 lines are rejected for classification. In addition, a running statistic for the number of scan lines/columns per 10 lines/columns is computed. If that value exceeds 3 then the 10 lines/columns for that statistic is rejected for classification.

3.5.5.3 Detection

We assume that the imagery is provided to the module in segments or scenes (as opposed to a continuous stream). Two statistics are computed. The first is the mean digital number for each scan line/column and the other is a 10 line/column sliding mean. If the mean for a scan line/column is less than a TBD threshold (e.g., 5), the line/column is considered suspect. Since a small digital number could be derived from a large contiguous region of water/ocean (in the VNIR or SWIR) or a large contiguous region of thick, very cold cloud (in the TIR), the second mean is utilized. If the difference between the line/column mean and the running mean is greater than a TBD threshold (e.g., 5), the line/column is classified as missing. If a coastal region or cold cloud boundary were perfectly aligned along a scan line/column, the second test (i.e., the difference in means) could incorrectly classify lines/columns as missing; however, this possibility of this condition is very remote.

3.5.5.4 Reconstruction

After all the bands have been screened for missing lines/columns, the lines/columns that pass the criteria of section 3.2.5.2 are reconstructed. Two types of reconstruction are required, either for 1 lone missing scan lines/column or 2 adjacent scan lines/columns. One lone missing scan line/column is linearly interpolated from the 2 adjacent scan lines. Two missing scan lines are replaced by their adjacent nearest neighbors.

An alternative technique is being investigated which should provide better reconstruction than the aforementioned technique. It is taken from Mather (1991). The technique requires that another highly cross correlated band be available and that the missing lines/columns being reconstructed in one band are not also missing in the cross correlated band. For example, if ASTER band 5 has a missing column then band 6 is used in the reconstruction. Similarly, if band 10 has a missing line than band 11 is used in the reconstruction. A digital number v of a specific pixel in a single missing line is computed as follows:

$$v_{i,j,k} = M \{v_{i,j,r} - (v_{i,j+1,r} + v_{i,j-1,r}) /2\} + (v_{i,j+1,k} + v_{i,j-1,k}) /2$$

where i and j are the column and line indices, respectively, and M is the ratio of the standard deviation of the pixels in line j of band k to the pixels in line j or band r. For a missing column, the equation is modified as follows:

$$v_{i,j,k} = M \{v_{i,j,r} - (v_{i+1,j,r} + v_{i,-1,j,r}) /2\} + (v_{i+1,j,k} + v_{i-1,j,k}) /2$$

For reconstructing pairs of missing lines/columns the indices j-1, j+1, i-1, and i+1 are replaced in band k by j-2, j+2, i-2, and i+2 as appropriate.

3.5.5.5 Striped Imagery

The module for destriping imagery will be applied in sequence following the missing scan line/column module as a preprocessing step to generating the cloud mask. The destriping module, like the missing scan line/column module, is performed in 2 steps. First is detection of striping and the second, if necessary, is reconstruction. We assume that since the VNIR and SWIR are pushbroom sensors that imagery from these sensors do not suffer from striping. However, since the TIR is a scanning sensor, striping can occur, and we assume that, if present, it has a periodicity equal to the number of scan elements in the sensor (e.g., 10 lines).

3.5.5.6 Detection

As striping is periodic, the frequency domain is used for detection. In each of the 5 TIR bands 15 columns from the approximately 700 columns of imagery are Fourier Transformed. The 15 columns of data are selected uniformly across the width of the imagery. If the average magnitude spectrum for these 15 columns exceeds a TBD threshold near the spatial frequency corresponding to expected striping periodicity (i.e., 1 cycle per 20 lines) and the first harmonic (i.e., 2 cycles per 20 lines), the entire image is flagged as striped.

3.5.5.7 Reconstruction

If the imagery is flagged as striped, then a technique from Mather (1991) is applied to the imagery. As in section 3.2.5.4, we assume that the imagery is provided as a segment or scene. The technique is based on histogram matching and assumes that the relative frequency of imaged feature (i.e., actual radiances) do not vary from one element of the scan sensor to another across the entire scene. The technique is implemented as follows. A target or nominal cumulative histogram (of gray levels) is computed for the entire image. A set of cumulative histograms (i.e., 10) then is computed for each of the lines corresponding to a specific sensor element. For example, if there are 10 elements in the TIR scan sensor then 10 cumulative histograms are computed, one each for lines 1, 11, 21, ... and lines 2, 12, 22, ..., etc. A table lookup is then constructed for each possible gray level (0 to 4095) and each of the 10 sensor elements. The table lookup values are derived from histogram matching each of the 10 element cumulative histograms to the target or entire image cumulative histogram. For example, the frequency value for a specific gray level in an element histogram is compared to the target histogram. The gray level that corresponds to the first frequency value in the target histogram greater than the frequency value in the element histogram then becomes the new gray level in the destriped imagery (i.e., the table lookup value). All of the pixels in the scene are then remapped to a new set of gray levels derived from the table lookup values.

Another technique is being investigated which is conceptually simpler; however, implementation is pending testing. The technique is based on the work of Pan and Chang, 1992. In this technique a specially designed finite impulse response (FIR) bandpass filter is convolved with the columns in the imagery. The filter is designed to attenuate only the spatial frequencies corresponding to the expected periodicity of the striping and at least one harmonic. The filter is approximately 15 pixels in length and is slid over the entire column of data.

3.5.5.8 Missing Bands

ASTER Channels		If Not Available Then We Use	
(1)	0.56 μm	(2)	0.66 μm
(3n)	0.81 μm	(2)	0.66 μm
(4)	1.65 μm		
(5)	2.16 μm	(6)	2.20 μm
(10)	8.3 μm	(11)	8.65 μm
(13)	10.6 μm	(12)	9.1 μm
(14)	11.3 μm		

4.0 CONSTRAINTS, LIMITATIONS, ASSUMPTIONS

The most significant constraint is that the current algorithm is developed only for daytime data acquisitions. The current daytime algorithm will only be exercised on data in which the solar zenith angle is less than 85 degrees. The nighttime algorithm is in the process of extensive development.

Inadequate numbers of LANDSAT TM scenes have been processed at present. We cannot guarantee and do not believe that the entire range of representative samples has been investigated. In particular, Arctic TM scenes with different cloud types, land, and tundra are required. The MAS data acquired for the Beaufort Sea will partially alleviate this problem. Additional LANDSAT Polar scenes are being acquired.

For the present algorithm development, the LANDSAT TM calibration is adequate for channels 1-5 and 7. However, the infrared channel 6 is essentially uncalibrated. This has caused difficulties in our algorithm development. For example, ocean surfaces often are retrieved with surface temperatures on the order of -12°C. However, unless they are covered with ice, the ocean temperatures cannot be below -1.8°C. Our solution to this calibration problem has been to examine the other visible/near-infrared channels for indications of ice. Where none are found, we arbitrarily rescale the channel 6 radiances such that open ocean water (away from any identifiable sea ice) is at -1.8°C. Another problem is that most of the present 24 LANDSAT scenes do not have water in the scene.

We are forced to take the average of these rescaling values found in the water scenes and to apply it to the non-water scenes.

Due to the very high spatial resolution of the LANDSAT TM imagery, and using the IVICS image display system with its capability for multiple channel overlays, channel differences and ratios, and 3-D clustering, we are relatively confident that most features can be identified accurately for the daytime scenes. The most difficult case is thin cirrus over a featureless snow-covered background. Availability of MAS and TIMS data will allow us to validate the nighttime algorithm.

REFERENCES

Allen, A. G., R. M. Hanson, J. W. Erismen, 1989: Field measurement of the dissociation of ammonium nitrate and ammonium chloride aerosols. *Atmos. Env.*, **23**, 1591-1600.

Barry, R. G., A. Henderson-Sellers, and K. P. Shine, 1984: Climate sensitivity and the marginal crysosphere. In *Climate Processes and Climate Sensitivity (Geophysics Monograph 29)*, ed. J. E. Hansen and T. Takahashi, Am. Geophys. Union. 221 pp.

Coakley, J. A. and F. P. Bretherton, 1982: Cloud cover from high-resolution scanner data: Detecting and allowing for partially filled fields of view. *J. Geophy. Res.*, **87**, 4917-4932.

Cox, E., 1992: Fuzzy Fundamentals, *IEEE Spectrum*, October 1992, 58-61.

Crane, R. G. and M. R. Anderson, 1984: Satellite discrimination of snow/cloud surfaces. *Intnl. J. Rem. Sens.*, **5**, 213-221.

Duda and Hart, 1973: Pattern Classification and Scene Analysis, John Wiley & Sons.

Ebert, E., 1987: A pattern recognition technique for distinguishing surface and cloud types in the polar regions, *J. Clim. Appl. Meteor.*, **26**, 1412-1427.

Ebert, E., 1989: Analysis of polar clouds from satellite imagery using pattern recognition and a statistical cloud analysis scheme, *J. Appl. Meteor.*, **28**, 382-399.

Ebert, E., 1992: Pattern recognition analysis of polar clouds during summer and winter. *Int. J. Remote Sensing*, **13**, 97-109.

Feind, R. E. and R. M. Welch, 1994: Registering TIMS to AVIRIS Imagery using Cloud Morphology. Submitted for Review.

Fuji and Morita, 1971: Recognition systems for handwritten simulating visual nervous systems. Pattern Recognition and Machine Learning, pp 56-69, Plenum Press, New York.

Giarratano J. and G. Riley, 1989: *Expert Systems: Principle+s and Programming*, 632 pp., PWS-KENT Publishing Company, Boston, 1989.

Haralick, R. M., K. S. Shammugam, and I. Dinstein, 1973: Textural features for image classification. *IEEE Trans. On Systems, Man, and Cybernetics*, **Vol. SMC-3**, No. 6, 610-621.

Hecht-Nielsen, R., 1990: *Neurocomputing*. Addison-Wesley, Reading, MA, 430 pp.

Key, J., 1990: Cloud cover analysis with Arctic AVHRR data. Part II: Classification with spectral and textural measures. *J. Geophys. Res.*, **95**, 7661-7675.

Key, J. and R. G. Barry, 1989: Cloud cover analysis with Arctic AVHRR data. 1. Cloud detection. *J. Geophys. Res.*, **94**, 8521-8535.

King, M. D. and S. C. Tsay, 1993: Theoretical basis of cloud retrieval algorithms for MODIS: Cloud cover, thermodynamic phase, optical thickness, and effective particle radius. MODIS Algorithm Theoretical Basis Document, NASA Goddard Space Flight Center.

Kuo, K-S, R. M. Welch, and S. K. Sengupta, 1988: Structural and textural characteristics of cirrus clouds observed using high spatial resolution LANDSAT imagery, *J. Appl. Meteor.*, **27**, 1242-1260.

Lee, J., R. Weger, S. K. Sengupta, and R. M. Welch, 1990: A neural network approach to cloud classification. *IEEE Trans. Geosci. and Remote Sens.*, **28**, 846-855.

Li, Z. and H. G. Leighton, 1991: Scene identification and its effect on cloud radiative forcing in the Arctic. *J. Geophys. Res.*, **96**, 9175-9188.

Logar, A., E. M. Corwin, and W. J. Oldham, 1992: Predicting acid concentrations in processing plant efficient: An application of time series prediction using neural networks. In Proc. ACM SIGAPP Symposium or Applied Computing, 651-658.

Logar, A., E. M. Corwin, and W. J. Oldham, 1994: A performance comparison of classification techniques for multi-font character recognition. To appear International Journal of Man Machine Studies.

Logar, A., E. M. Corwin, Watters, R. C. Weger, and R. M. Welch, 1994: A don't care back propagation algorithm applied to satellite image recognition. Proceedings of National SCA/ACM Conference, March 1994.

Luger, G. F. and W. A., Stubblefield, 1989: *Artificial Intelligence and the Design of Expert Systems*, 600 pp., The Benjamin Cummings Publishing Company, Inc., 1989.

Mather P. M., 1991: *Computer Processing of Remotely-Sensed Images*. John Wiley and Sons, Copyright 1987.

McGuffie, K., R. G. Barry, A. Schweiger, D. A. Robinson, and J. Newell, 1988: Intercomparison of satellite derived cloud analyses for the Arctic ocean in Spring and Summer, *Inter. J. Rem. Sens.*, 9, 447-467.

Menzel, P., 1994: Cloud top properties and lcoud phase algrithm theoretical basis document. MODIS Team Document.

Menzel, P. and K. Strabala, 1993: Cloud top properties and cloud phase algorithm theoretical basis document. Cooperative Institute for Meteorological Satellite Studies, Space Science and Engineering Center, University of Wisconsin-Madison.

Minnis, P., and E. F. Harrison, 1984: Diurnal variability of regional cloud surface radiative parameters derived from GOES data: Part I Analysis Method; Part II: Cloud results; Part III: Radiative budget. *J. Climate Appl. Meteor.*, 23 (in press).

Nair, U. S. and R. M. Welch, 1994: Textural stability and sampling. [Submitted to *Intnl. J. Remote Sensing*]

Ormsby, J. P. and D. D. Hall, 1991: Spectral properties of fog over the Malospina Glacier, Alaska, in comparison to snow, ice, and clouds. *Photogrammetric Engineering and Remote Sensing*, **Vol. 57, No. 2**, 179-185.

Parkinson, C. L., J. C. Comiso, H. J. Zwally, D. J. Cavalieri, P. Gloersen, and W. J. Campbell, 1987: Arctic Sea Ice, 1973-1976: Satellite Passive Microwave Observations, NASA SP-489, Scientific and Technical Branch, Washington, DC. 296 pp.

Rabindra, P, S. K. Sengupta, and R. M. Welch, 1992: An interactive hybrid expert system for polar cloud and surface classification. *EnvironMetrics*, 3(2), 121-147.

Raschke, E., P. Bawer, and H. J. Lutz, 1992: Remote sensing of clouds and surface radiation budget over polar regions. *Int. J. Remote Sensing*, **13**, 13-22.

Rossow, W. B., 1989: Measuring cloud properties from space. A review. *J. Climate*, **2**, 201-213.

Rossow, W. B., L. C. Garder, and A. A. Lacis, 1989a: Global, seasonal cloud variations from satellite radiance measurements. Part I: Sensitivity of analysis, *J. Climate*, **2**, 419-458.

Rossow, W. B., C. L. Brest, and L. C. Garder, 1989b: Global, seasonal surface variations from satellite radiance measurements. *J. Climate*, **2**, 214-247.

Rumelhart, D. E., G. Hinton, and R. Williams, 1986: Learning internal representations by error propagation. In *Parallel Distributed Processing: Exploration in the Microstructure of Cognition.* D. Rumelhart and J. McClelland (eds.), MIT Press, pp. 318-362.

Sakellariou, N. K, and H. Leighton, 1988: Identification of cloud-free pixels in inhomogeneous surfaces from AVHRR radiances. *Journal of Geophysical Research*, **93**, D5, 5287-5293.

Saunders, and Kriebel, 1988: An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. J. Remote Sensing*, **9**, 123-150.

Schlesinger, M. E. and J. F. B. Mitchell, 1987: Climate model simulations of the equilibrium climatic response to increase carbon dioxide. *Rev. Geophys.*, **25**, 760-798.

Serra, J., 1982: *Image Analysis and Mathematical Morphology.* Academic Press, London. 610 pp.

Sèze, G., and W. B. Rossow, 1991a: Time-cumulated visible and infrared radiance histograms used as descriptors of surface and cloud variations. *Int. J. Remote Sens.*, **12**, 921-952.

Sèze, G., and W. B. Rossow, 1991b: Effects of satellite data resolution on measuring the space-time variations of surfaces and clouds. *Int. J. Remote Sens.*, **12**, 130-132.

Shine, K. P. and R. G. Crane, 1984: The sensitivity of one-dimensional thermodynamic sea ice model to a change in cloudiness. *J. Geophys. Res.*, **C89**, 10615-10622.

Simpson, J. J. and C. Humphrey, 1990: an automated cloud screening algorithm for daytime AVHRR imagery. *J. Geophys. Res.*, **C8**, 13,459-13,481.

Steffen, K. and J. E. Lewis, 1988: Surface temperatures and sea ice typing for northern Baffin Bay. *Intnl. J. Rem. Sens.*, **9**, 409-422.

Stowe, L. L., H. Y. M. Yeh, T. F. Eck, C. G. Wellemeyer, H. L. Kyle, and the NIMBUS-7 Cloud Data Processing Team, 1989: Nimbus-7 global cloud climatology, Part II: First year results. *J. Climate*, **2**, 671-709.

Tovinkere, V. R., 1992: *Fuzzy Logic Expert System for Classification of Polar Regions*, M.S. Thesis, Department of Mathematics and Computer Science, South Dakota School of Mines and Technology, Rapid City, SD, 69 pp.

Tovinkere, V. R., M. Penaloza, A. Logar, J. Lee, R. C. Weger, T. A. Berendes, and R. M. Welch, 1993: An intercomparison of artificial intelligence approaches for polar scene identification. *J. Geophys. Research*, **98**(D3), 5001-5016.

Watters, Logar, and Corwin, 1993: A comparison of neural network algorithms for polar scene classification. Artificial Neural Networks in Engineering.

Welch, R. M., K. S. Kuo, S. L. Sengupta, 1990: Cloud and surface textural features in polar regions. *IEEE Trans. Geosci, and Remote Sens.*, **28**, 520-528.

Welch, R. M., S. K. Sengupta, and D. W. Chen, 1990: Cloud field classification based upon high spatial resolution textural features. 1. Gray level co-occurrence matrix approach. *J. Geophys. Res.*, **93**, 12663-12681.

Welch, R. M., S. K. Sengupta, and K. S. Kuo, 1988: Marine stratocumulus cloud fields off the coast of southern California observed using Landsat imagery. Part II: Textural analysis. *J. Appl. Meteor.*, **27**, 362-378.

Welch, R. M., S. K. Sengupta, A. K. Gorock, P. Rabindra, N. Rangaraj, and M. S. Navar, 1992: Polar cloud and surface classification using AVHRR imagery: An intercomparison of methods, J. Appl. Meteor., **31**, 405-420.

Wetherald, R. T. and S. Manabe, 1980: Cloud cover and climate sensitivity. *J. Atmos. Sci.*, **37**, 1485-1510.

Wiscombe, W. J. and S. G. Warre, 1981: A model for the spectral albedo of snow. I. Pure snow. *J. Atmos. Sci.*, **37**, 2712-2727.

Yamanouchi, T. and S. Kawaguchi, 1992: Cloud distribution in the Antarctic from AVHRR data and radiation measurements at the surface. *Int. J. Remote Sensing*, **13**, 111-127.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	11 July 1994	Semi-annual, 1 Jan 94-30 Jun 94

4. TITLE AND SUBTITLE
The Effect of Cloud Inhomogeneities Upon Radiative Fluxes, and the Supply of a Cloud Truth Validation Dataset

5. FUNDING NUMBERS
C - NAS5-31718

6. AUTHOR(S)
Ronald M. Welch

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute of Atmospheric Sciences
S.D. School of Mines and Technology
501 E. St. Joseph Street
Rapid City, SD 57701-3995

8. PERFORMING ORGANIZATION REPORT NUMBER
SA-5

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt Road
Greenbelt, MD 20771

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The polar region cloud mask is being developed in support of the Aster Science Team. We have 68 Landsat TM polar scenes (24 from the Antarctic and 44 from the Arctic) which are being used as the basis for algorithm development and testing. A wide range of spectral and textural features and numerous algorithms are being examined to determine accuracy and CPU requirements. The aim is to optimize the algorithm. The development of algorithms for destriping and replacing missing lines in imagery is also in progress.

Monte Carlo simulations of 3-D clouds represented by hyperboloids of 2 sheets is now in progress. A paper detailing the registration of TIMS to AVIRIS using cloud morphology has been submitted to IEEE Geoscience and Remote Sensing. An analysis of cloud and background properties also appears in that paper.

14. SUBJECT TERMS	15. NUMBER OF PAGES
ASTER, polar scene classification, Monte Carlo, TIMS, AVIRIS, registration	71
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.